International Edition

The Economics of Health and Health Care

Sherman Folland
Allen C. Goodman
Miron Stano

ROUTLEDGE

The Economics of Health and Health Care

8th Edition

Folland, Goodman, and Stano's bestselling *The Economics of Health and Health Care* text offers the market-leading overview of all aspects of Health Economics, teaching through core economic themes, rather than concepts unique to the health care economy.

The Eighth Edition of this key textbook has been revised and updated throughout, and reflects changes since the implementation of the Affordable Care Act (ACA). In addition to its revised treatment of health insurance, the text also introduces the key literature on social capital as it applies to individual and public health, as well as looking at public health initiatives relating to population health and economic equity, and comparing numerous policies across Western countries, China, and the developing world. It provides up-to-date discussions on current issues, as well as a comprehensive bibliography with over 1,100 references. Extra material and teaching resources are now also available through the brand new companion website, which provides full sets of discussion questions, exercises, presentation slides, and a test bank.

This book demonstrates the multiplicity of ways in which economists analyze the health care system, and is suitable for courses in Health Economics, Health Policy/Systems, or Public Health, taken by health services students or practitioners.

Sherman Folland is Professor of Economics (emeritus) at Oakland University in Rochester, Michigan, USA. He received his PhD from The University of Iowa in 1975 and has taught at Penn State University and Oakland University. He has written over 50 articles and edited several books.

Allen C. Goodman is Professor of Economics at Wayne State University in Detroit, Michigan, USA. He received his PhD from Yale University and has taught at Lawrence University, The Johns Hopkins University, and Wayne State University. He has written over 100 articles and three books in health and housing economics.

Miron Stano is Professor of Economics and Management at Oakland University in Rochester, Michigan, USA. He received his PhD from Cornell University in 1971 and has taught at Wayne State University and Oakland University. He has written over 70 articles and has also served as an expert witness in antitrust cases involving the health care professions.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

The Economics of Health and Health Care

International Edition

8th Edition

Sherman Folland, Allen C. Goodman, and Miron Stano

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

Eighth edition published 2017
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

First edition published by Macmillan 1993
Seventh edition published by Pearson 2013

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-138-20805-6 (pbk)
ISBN: 978-1-315-10348-8 (ebk)

Typeset in Sabon
by Apex CoVantage, LLC

Visit the companion website: www.routledge.com/cw/folland

Contents

Preface

This is the eighth edition of a book conceived in 1990 and first published in 1993. The world of health economics has changed since 1990. Our first (1993) edition made but a single reference to "managed care." Until the seventh edition, we spoke of the United States as the only country without a universal health insurance plan. This, of course, changed with the passage of the Affordable Care Act (ACA) in March 2010.

New to the Eighth Edition

Five years have passed since our last edition, an eternity in the health care world. Outside of updated tables, and features, we highlight the following.

- The Affordable Care Act (ACA) withstood two legal challenges and changed the landscape of health care. By 2016, 20 million more Americans now have health insurance because of the ACA, and over 80 percent with a plan have government subsidies. The Congressional Budget Office estimates that were the ACA repealed, 24 million would have no health insurance. Although the Trump administration has pledged to "repeal and replace" Obamacare, preliminary (2017) analyses suggest that change will occur along the margins of the Act rather than wholesale repeal.
- A new Chapter 24 on Social Capital replaces our previous chapter on Epidemiology and Economics. Social capital is an exciting new development in the conceptualization and modeling of health care, health services, and population health.
- We have streamlined our material on regulation by combining the previous Chapters 19 and 20 into a new Chapter 19 entitled Government Intervention in Health Care Markets
- Chapter 20 (Social Insurance) continues with the most up-to-date evaluations of the Medicare Advantage, and the Medicare Part D drug benefit. It also updates research and policy work on the inherent conflicting incentives between the Medicare and Medicaid programs that jointly lead to inefficiencies in the provision of health services for the elderly and the poor.
- Chapter 21 continues our path-breaking comparative analyses across countries. We introduce a new classification of health care systems, supplanting the Gordon typology that we have used for over 20 years. We review our discussion of the the Chinese health economy which has moved away from the command system of the 1950s through the 1970s, toward a more incentive-based system, with its conflicting impacts. Chapter 21 also provides updated survey information on comparative satisfaction across eleven countries, looking in particular at differences in access and in costs.

From the beginning, we have sought to assist instructors in conveying a clear, step-by-step understanding of health economics to their students. We have also believed it important for instructors to demonstrate what health economics researchers are doing in theory and in empirical work. The book synthesizes contemporary developments around a set of economic principles including maximization of consumer utility and economic profit, and it makes these principles accessible to undergraduate as well as to graduate students. Rather than focusing on institutions specific to the health care economy, we continue to emphasize core economics themes as basic as supply and demand, as venerable as technology or labor issues, and as modern as the economics of information. We continue to improve accessibility to the book for the wide range of health services students and practitioners.

Students must have a working knowledge of the analytical tools of economics and econometrics to appreciate the field of health economics. Some students may be ready to plunge directly into Chapter 5, "The Production of Health," upon completion of the introductory Chapter 1. However, Chapters 2 through 4 help students and their instructors to develop or to review needed analytical concepts before tackling the core subject matter. In Chapter 2, students with as little as one semester of microeconomics may review and study how economists analyze problems, using relevant health economics examples. Chapter 3 provides a review of core statistical tools that characterize modern economic and health services analyses. Chapter 4 completes the core economic concepts by reviewing the concept of economic efficiency, and showing how cost-benefit, cost-effectiveness, and cost-utility analyses fit into the general economic framework.

Consistent with an emphasis on clear exposition, the book makes extensive use of graphs, tables, and charts. As in all previous editions, we require no calculus. Discussion questions and exercises help students master the basics and prompt them to think about the issues. We also include up-to-date applications of theory and policy developments as features, and occasional tidbits containing purely background information.

We caution that some chapters, such as those on insurance and on regulation, although developed without advanced mathematics, are logically complex and will require considerable effort. No painless way is available to appreciate the scope of the contributions that scholars have made in recent years. More advanced students of the health care economy who seek further challenges can utilize a comprehensive references section, with over 1,100 sources, to enrich their (and our) work through referral to the original sources.

Additional Sources

The Internet now contains tables and charts that were once available only in book form and then only after several years. We have chosen to focus on those sites that we believe to be both long lasting and reliable.

- Bureau of the Census, for health insurance (www.census.gov/newsroom/press-releases/2016/cb16-158.html)
- Centers for Disease Control and Prevention (www.cdc.gov)
- Centers for Medicare and Medicaid Services, for research, statistics, data, and systems (www.cms.hhs.gov/home/rsds.asp)
- Kaiser Family Foundation (www.kff.org), specializing in studies of health insurance
- National Institutes of Health (www.nih.gov)

- Organization for Economic Cooperation and Development, for international data (www.oecd.org)
- Social Security Administration, for research and analysis (www.ssa.gov/policy/research_subject.html)

Health-related material is increasingly accessible both in print and on the Internet. University-affiliated professionals and their students will discover that their libraries have extensive electronic access to a wide range of journals. Most health economists browse *Health Affairs*, an up-to-date policy journal. *Health Economics*, *Journal of Health Economics*, and *American Journal of Health Economics* have emerged as the leading technical journals that specialize in health economics. Users can see, from our comprehensive reference section, many other specialized journals, including health services and medical journals not often referenced by economists. In the popular press, the *New York Times* and the *Wall Street Journal* also provide excellent health economics coverage.

The Handbook of Health Economics, a two-volume set published in 2000, with an additional volume in 2011, emerged as an invaluable source for specific topics, with more detail and more mathematic rigor than any text, including this one. The *Elgar Companion to Health Economics*, published in 2006 and updated in 2012, provides both useful updates and important new topics. They are not texts, however. Our book, with its graphical analysis, discussion questions, and problem sets, provides a valuable complement to both *Handbook* and *Companion* offerings.

Alternative Course Designs

The economics of health and health care encompasses an evolving literature with no single "correct" order for the course design. U.S. economists typically organize topics through markets, and include the roles of government much later. International health economists, and population and public health students and scholars, often assign the governmental sector far more importance; it is "public" health, after all. No matter how we construct it, a text is necessarily linear in that one chapter must follow another.

Our text offers instructors considerable flexibility. We view the 24 chapters as six parts:

I. Basic Economics Tools (Chapters 1–4)
II. Supply and Demand (Chapters 5–9)
III. Information and Insurance Markets (Chapters 10–13)
IV. Key Players in the Health Care Sector (Chapters 14–17)
V. Social Insurance (Chapters 18–22)
VI. Special Topics (Chapters 23–24)

The categories are not exclusive. Chapter 8, looking at the demand and supply of insurance, is as important to Part III on insurance as it is to Part II on demand and supply of goods.

From front to back, we follow an "economics" model in which we first examine consumers and firms in a world without government and governmental policies. As a result, explicit discussions of government policies do not come until Chapter 19, although we examine regulation, licensing, and mandates in reference to other topics much earlier. Many economics instructors may wish to follow the chapters in the book's order.

Instructors with population health, public health, or policy interests may wish to "tool up" on some of the earlier analyses and then skip directly to Part V, which examines social insurance, health care regulation, and health care reform. After that, they may wish to browse selected topics. Some analyses build on one another within chapters, but we seek to minimize cross-referencing among chapters.

Instructor Resources

The Economics of Health and Health Care is connected to the Companion Website available at www.routledge.com/cw/folland. As a registered faculty member, you can download resource files. The following supplements are available to adopting instructors:

- Instructors' Answers, Teaching Tips, and Study Material.
- PowerPoint Slides.

The Publisher would like to thank Laura Storino for providing the PowerPoint Slides.

The International Handbook on Teaching and Learning Economics (2012) has a section by Allen Goodman on the teaching of health economics, along with over 70 chapters on general course content, specific fields, pedagogic techniques, and the scholarship of the teaching enterprise.

Acknowledgments

This book emerges from years of classroom and professional interaction, and we thank those students and colleagues who have discussed the economics of health and health care with us and who have challenged our ideas. Annie Todd suggested the endeavor back in 1989 and introduced us to each other, and Tom McGuire convinced us that we could succeed. We have been blessed with outstanding editors—Jill Leckta, Leah Jewell, Rod Banister, Gladys Soto, Marie McHale, Michael Dittamo, Chris Rogers, Susie Abraham, and Lindsey Sloan. We are delighted to work with Emily Kindleysides on this edition.

We have received outstanding composition support from our publishers, with John Travis, Maureen Wilson, Kelly Warsak, Melissa Feimer, Clara Bartunek, Kerri Tomasso, and Nancy Freihofer managing previous editions. Laura Johnson has been our most capable manager for this eighth edition.

We extend our gratitude to instructors at hundreds of universities in the United States and around the world, who have helped us pass the market test by adopting the book. Numerous professional colleagues generously offered their time and energy to read and critique various chapters. We thank:

Carson Bays, East Carolina University
James Burgess, Boston University
Frank Chaloupka, University of Illinois–Chicago
Hope Corman, Rider College
Merton Finkler, Lawrence University
Colleen Flood, University of Toronto
Howard Forman, Yale University
Andrew D. Foster, Brown University
Gary Fournier, Florida State University
H. E. Frech, III, University of California–Santa Barbara
Martin Gaynor, Carnegie Mellon University
John Goddeeris, Michigan State University
Warren Greenberg, late of George Washington University
Michael Grossman, City University of New York
Jack Hadley, Center for Health Policy Studies
James Hamilton, Wayne State University
Janet Hankin, Wayne State University
Govind Hariharan, Kennesaw State University
Ronald Horwitz, Oakland University
Harold Hotelling, late of Lawrence Technological University

Kathy A. Kelly, University of Texas at Arlington
Paul Langley, University of Colorado
William Low, University of Northern British Columbia
Timothy McBride, Washington University
Tom McCarthy, National Economic Research Associates
Catherine McLaughlin, University of Michigan
Eric Nauenberg, University of Toronto
Eleanor Nishiura, Wayne State University
John Nyman, University of Minnesota
Albert A. Okunade, University of Memphis
Martin Pernick, University of Michigan
Uwe E. Reinhardt, Princeton University
Joann Richards, Oakland University
John Rizzo, Stony Brook University
T. Paul Schultz, Yale University
Ya-chen Tina Shih, MEDTAP International, Inc.
Kosali Simon, Indiana University
Robert J. Sokol, Wayne State University
John M. Tilford, University of Arkansas
Scott Vander Linde, Calvin College
Rob Verner, Kent State University
Jian Wang, Peking University
Kenneth Warner, University of Michigan
Gerard Wedig, University of Pennsylvania
Burton Weisbrod, Northwestern University
William White, Cornell University
Barbara Wolfe, University of Wisconsin–Madison
Michael Woolley, University of Southern California

We also thank the following colleagues who, with their classes, test-taught and test-learned parts of the text:

Randall Ellis, Boston University
Gail Jensen, Wayne State University
Thomas McGuire, Harvard University
Lee R. Mobley, Georgia State University
Kathleen Possai, Wayne State University
Richard Scheffler, University of California–Berkeley
Nancy Wolff, Rutgers University

In the 27 years since we began this project, our wives have patiently supported our efforts and our young daughters have grown into talented young professional women. All have watched and helped their husbands and fathers learn about health economics. For these gifts we remain eternally grateful.

S. F.
Rochester, Michigan
A. C. G.
Detroit, Michigan
M. S.
Rochester, Michigan

Chapter 1

Introduction

In this chapter

- What Is Health Economics?
- The Relevance of Health Economics
- Economic Methods and Examples of Analysis
- Two Notable Contributors to Health Economics
- Does Economics Apply to Health and Health Care?
- Is Health Care Different?
- Conclusions

Health care accounts for over one-sixth of the U.S. economy! Yet millions in the United States have no health insurance. The Obama reform legislation, the Affordable Care Act or ACA, passed in 2010 and still in progress, is designed to address this problem. Health, health care costs, and health insurance have dominated the economic and political landscape in the United States and many other countries. Health economists study these issues carefully, and in general, health economics applies the principles of scientific empiricism to issues of health and health care. Because our health is of vital concern to us, and because the health care sector has become the largest sector of the U.S. economy, we should not be surprised that health economics has emerged as a distinct specialty within economics.

The scope of health economics and the emphasis of this text can be previewed by examining the Table of Contents. The production of health and health care, demand and supply of specific health services are prominent. Private health insurance markets critically define the U.S. workplace, so we examine these markets. Government, through its social programs and power to regulate, receives close attention. We also concentrate on issues such as information, quality of care, and equity of access. Finally, we look to the health care systems of other countries for information on their practices and for potential insights on the policy issues that dominate the political landscape.

In this first chapter, we provide further background on health economics and health economists. We follow with a broad overview of the magnitude and importance of the health care sector and with an introduction to some major policy concerns. As our final goal, we seek to promote the theme that economics helps explain how health care markets function. We focus on methods used in economic analyses and address two recurring questions: Is health care different, and does economics apply? Despite stressing the distinctive features of health care services and markets, we answer both in the affirmative. With appropriate modifications to conventional analytical tools, economics is relevant and useful. As we shall see throughout the book, although there is continuing controversy on many major issues, health economists have provided insight and solutions to most problems of academic and policy interest.

What Is Health Economics?

Health economics is defined by *who* health economists are, and *what* they do! Morrisey and Cawley (2008) examined the field of health economics in 2005 and found that almost all (96 percent) held academic doctorate degrees. Nearly three-quarters of those with doctorates received their degrees in economics.

The majority worked in university settings; most others worked for nonprofit organizations or in government, mainly the federal government. Health economists held their appointments in economics departments, schools of public health, and in schools of medicine. Many of the leading economics departments—e.g., MIT, Princeton, Berkeley, Harvard—now feature prominent health economists. Health economists draw on various sub-disciplines of training within economics, including labor economics, industrial organization, public finance, cost-benefit analyses, and most generally, microeconomics.

Throughout this book, we describe many specific research studies. Consider, at this time, that the United States devotes by far the largest share of GDP to health care spending (over one-sixth), and its per capita health care spending (over $9,500) greatly exceeds that of any other country. Most health economists agree that these spending patterns reflect the rapid rate of adoption of new technology in the United States. The United States does not have a very impressive record in terms of broad health outcomes indicators such as life expectancy

and infant mortality. Critics of the U.S. health care system often wonder what Americans are getting for their money. Policymakers and health economists seek to determine whether spending on new technology is worth it. Arguably, there is no more important issue.

Consider, for example, a new surgical procedure for a patient with acute myocardial infarction (heart attack). It is not enough to estimate the immediate cost impact of the new procedure and the expected benefit to the patient in terms of short-term survival. By impacting the patient's health for many years, the new treatment will affect spending well into the future. David Cutler (2007) develops a framework to address these complex interrelationships in "The Lifetime Costs and Benefits of Medical Technology." He analyzes *revascularization*, a set of surgical procedures such as coronary bypass and angioplasty that restore blood flow. He looks at a group of Medicare patients who have had heart attacks and he tracks them for up to 17 years. Chapter 4 devotes considerable attention to Cutler's work, but here we highlight his conclusion that revascularization costs $33,000 for an extra year of life. Is this worth it? Most would agree that it is!

Health care costs in general, and technology-related costs in particular, are relevant to all countries (Box 1.1 provides an international perspective). Health economics is still a relatively new discipline with an evolving scope and pedagogy, and neither it, nor we, will provide answers to all the health system questions that nations face. Despite this caveat, we cannot think of any field of study that is more relevant to unraveling the meaning of today's headlines, or more pertinent to the lives of individuals.

BOX 1.1

Technological Change and Health Care Costs—Why Rising Health Care Costs Affect All Nations

In a March 2005 speech to the National Association of Business Economics, then-Chair of the Council of Economic Advisers Harvey Rosen noted that over the last several decades, the health care quality—diagnostic techniques, surgical procedures, and therapies for a wide range of medical problems—has improved. Treatment of a heart attack today is simply not the same "commodity" as treatment of a heart attack in 1970. Although innovations like coronary bypass surgery and cardiac catheterization have raised expenditures per heart patient, they have actually reduced the prices of obtaining various health outcomes, such as surviving hospitalization due to a heart attack.

Some improvements in medical technique were quite inexpensive. Prescribing aspirin for heart attack victims leads to a substantial improvement in their survival probabilities, but new medical technologies were often costly. For example, it cost about $2 million to acquire a PET (positron emission tomography) machine, which can detect changes in cells before they form a tumor large enough to be spotted by X-rays or MRI. Such costly improvements lead medical expenditures to grow.

This technology-based theory also helps explain why countries as different as the United States, the United Kingdom, or Japan have all experienced increases in health care expenditures. Rosen argued that these societies have at least one thing in common—they all have access to the same expensive innovations in technology. The technology-based explanation puts any debate over cost containment in a new

light. Is it a bad thing if costs are rising mostly because of quality improvements? A key question in this context is whether people value these innovations at their incremental social cost. No one knows for sure, but economist Dana Goldman reiterates a provocative insight: "If you had the choice between buying 1960s medicine at 1960s prices or today's medicine at today's prices, which would you prefer?" A vote for today's medicine is validation of the improvement and willingness to pay for improved quality!

Source: Dana P. Goldman, "Pressure from Rising Health-Care Costs: How Can Consumers Get Relief?" www.rand.org/commentary/102305PE.html, accessed November 2016.

The Relevance of Health Economics

The study of health economics is important and interesting in three related ways: (1) the size of the contribution of the health sector to the overall economy, (2) the national policy concerns resulting from the importance many people attach to the economic problems they face in pursuing and maintaining their health, and (3) the many health issues that have a substantial economic element.

The Size and Scope of the Health Economy

The health economy merits attention for its sheer size, constituting a large share of GDP in the United States, as well as in other countries. It also represents a substantial capital investment and a large and growing share of the labor force.

Health Care's Share of GDP in the United States

By the second decade of the twenty-first century, more than $1 out of every $6 spent on final goods and services in the U.S. economy went to the health sector. As recently as 1980, the share of GDP (the market value of final goods and services produced within the borders of a country in a year) devoted to health care was $1 in $11, and in 1960 it was just $1 in $20. Figure 1.1 tracks the health economy's share of GDP from 1970 to 2024. The conclusion? The health care sector is a large and growing portion of our economy.

In calculating the share of GDP spent on health care, we net out the effects of general inflation. Therefore, only three major possibilities exist to explain the substantially increased ratios shown in Figure 1.1:

1 People may be buying more health services. Patients may be consulting with health care providers more frequently, doctors may be ordering more tests, or they may be prescribing more drugs.
2 People may be buying higher-quality health services, including products and services that previously were not available. Laser surgery, organ transplants, measles vaccines, and new treatments for burn victims, unavailable in 1960, have raised the quality of care. Economic theory suggests that people are willing to pay more for better quality.
3 Health care inflation may be higher than the general inflation rate. Higher incomes and the increased prevalence of insurance, including large government programs such as Medicare and Medicaid, may have led to increased health care prices over time.

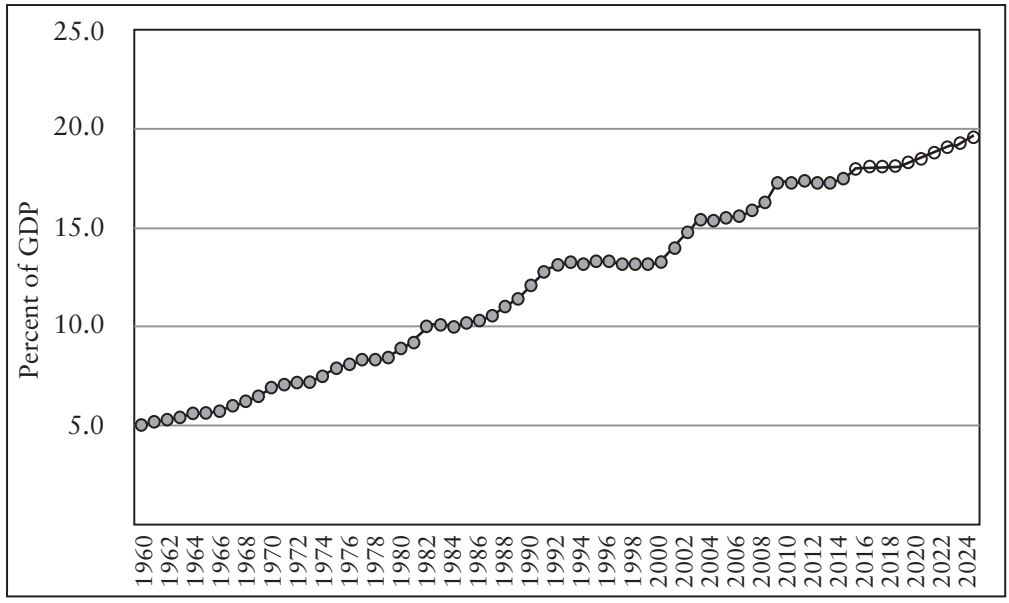


Figure 1.1 U.S. Health Expenditure Shares, 1960–2024

Sources: Centers for Medicare and Medicaid Services: www.cms.gov/NationalHealthExpendData, accessed November 15, 2016; Proquest Statistical Abstract of the United States, 2016; NHE figures from 2016 and later are projected numbers.

We seek to understand these phenomena and their contributions to total spending. The study of demand, insurance, production, technology, and labor supply, among other topics, will help meet this challenge.

Health Care Spending in Other Countries

Examining the health economies of other countries enhances our understanding of the U.S. health economy. Many countries have large health care sectors and face the same major issues. Table 1.1 shows how health care spending as a share of GDP grew rapidly in most countries between 1960 and 1980. A more mixed picture emerges after 1980. The health care share in the United States continued to grow in each period after 1980 shown in Table 1.1, but growth was more modest in most other countries.

The data also indicate the relative size of the U.S. health economy compared to that of other countries. For example, health care's share of GDP in the United States is nearly twice as large as the share in the United Kingdom—a country with national health insurance. Is care costlier in the United States? Is it higher quality care, or are we simply consuming more?

Importance of the Health Economy in Personal Spending

Because it accounts for such a large share of the domestic product, the size of the health economy is also reflected through other key indicators. Two of these are especially easy to

Table 1.1 Health Expenditures as Percent of GDP in Selected OECD Countries

Country	*1960*	*1970*	*1980*	*1990*	*2000*	*2010*	*2015*
Australia			6.3	6.9	7.6	8.5	9.3
Austria	4.3	5.2	7.5	8.4	9.2	10.1	10.3
Belgium	3.9	6.3	7.2	8.6	7.9	9.9	10.4
Canada	5.4	6.9	7.0	8.9	8.3	10.7	10.2
Czech Republic				4.7	5.7	6.9	7.6
Denmark			8.9	8.3	8.1	10.4	10.6
Finland	3.8	5.5	6.3	7.7	6.9	8.9	9.6
France	3.8	5.4	7.0	8.4	9.5	10.7	11.0
Germany		6.0	8.4	8.3	9.8	11.0	11.1
Greece		5.4	5.9	6.6	7.2	9.9	8.2
Hungary					6.8	7.6	7.0
Iceland	3.0	4.7	6.3	7.8	9.0	8.8	8.8
Ireland	3.7	5.1	8.3	6.1	5.9	10.6	9.4
Italy				7.7	7.6	9.0	9.1
Japan	3.0	4.6	6.5	6.0	7.4	9.5	11.2
Korea			3.4	4.0	4.0	6.4	7.2
Luxembourg		3.1	5.2	5.4	5.9	7.1	7.2
Mexico				4.8	4.9	6.2	5.9
Netherlands			7.4	8.0	7.1	10.4	10.8
New Zealand		5.2	5.9	6.9	7.5	9.7	9.4
Norway	2.9	4.4	7.0	7.6	7.7	8.9	9.9
Poland				4.8	5.3	6.4	6.3
Portugal		2.5	5.3	5.9	8.4	9.8	8.9
Slovak Republic					5.3	7.8	7.0
Spain	1.5	3.5	5.3	6.5	6.8	9.0	9.0
Sweden		6.8	8.9	8.2	7.4	8.5	11.1
Switzerland	4.9	5.4	7.3	8.2	9.3	10.5	11.5
Turkey			3.3	3.6	4.7	5.3	5.2
United Kingdom	3.9	4.5	5.6	6.0	6.3	8.5	9.8
United States	5.1	7.0	8.7	11.9	12.5	16.4	16.9

Note: OECD data for the United States may differ slightly from values reported by the Centers for Medicare and Medicaid Services.

Source: Organization for Economic Cooperation and Development (OECD) Health Care Data, extracted June 2016.

relate to at the personal level: (1) share of income spent on medical care and (2) number of jobs in the health economy.

Table 1.2 provides data on how U.S. consumers spend their disposable incomes. It shows that in 2015, consumers spent 18.3 percent of their budgets on health care, as opposed to 7.3 percent on food, and 15.6 percent on housing. These figures represent a major shift in spending patterns. As recently as 1960, food represented about 25 percent of spending, and medical care only 5 percent.

Importance of Labor and Capital in the Health Economy

Table 1.3 provides information on specific health care occupations and their growth since 1970. In 2013, there were over 1,045,910 physicians and almost 287,420 pharmacists. The

Table 1.2 Total Consumption Expenditures (in $ Billions) by Type, 2015

	2015	*% of Total*
Total personal consumption expenditures	**12,283.7**	**100.0%**
Durable goods	**1,355.2**	**11.0%**
Nondurable goods	**2,656.9**	**21.6%**
Food and beverages	900.7	7.3%
Clothing and footwear	379.5	3.1%
Gasoline and other energy goods	303.7	2.5%
Other nondurable goods	1,073.0	8.7%
Services	**8,271.6**	**67.3%**
Housing	1,919.9	15.6%
Household utilities	313.3	2.6%
Transportation services	368.4	3.0%
Recreation services	466.3	3.8%
Food services and accommodations	808.8	6.6%
Other services	2,147.2	17.5%
Health care	2,247.7	18.3%
Physicians	484.5	3.9%
Dentists	117.8	1.0%
Paramedical services	328.5	2.7%
Hospitals and nursing homes	1,138.2	9.3%
Health insurance	178.7	1.5%

Source: U.S. Department of Commerce, Bureau of Economic Analysis, Table 2.4.5: Personal Consumption Expenditures by Type of Product [Billions of Dollars], last revised on August 3, 2016, accessed August 2016. www.bea.gov/iTable/iTable.cfm?ReqID=9&step=1#reqid=9&step=3&isuri=1&903=70.

Table 1.3 Active Health Personnel and Number per 100,000 Population (in Parentheses)

Occupation	*1970 Total*	*1980 Total*	*1990 Total*	*2000 Total*	*2013 Total*
Physicians	334,028 (164)	467,679 (206)	615,421 (247)	813,770 (289)	1,045,910(330)
Licensed Practical and Licensed Vocational Nurses	—	—	—	679,470 (241)	705,200 (223)
Registered Nurses	750,000 (369)	1,272,900 (562)	1,789,600 (720)	2,189,670 (778)	2,661,890 (841)
Pharmacists	112,750 (55)	142,780 (63)	161,900 (65)	212,660 (76)	287,420 (91)
U.S. Population	203,302,031	226,542,199	248,709,873	281,421,906	316,799,000

Source: U.S. Department of Health and Human Services, *Health United States*, Various Years.

nursing sector alone consisted of over 3 million people with over three-quarters of them trained as registered nurses.

The considerable growth in health care personnel is evident. In 1970, there were 334,000 physicians, or 164 physicians per 100,000 people. By 2013, the number of physicians had increased by 171 percent to 1,045,910 or 330 per 100,000 population. The number of registered nurses had more than tripled by 2013, with their number per 100,000 population more than doubling from 369 to 841.

Reflecting the increases in spending, the health care sector serves increasingly as a source of employment. Thus, cutbacks in spending on health care, if proposed and implemented, would typically mean cutbacks in employment opportunities.

In addition to labor, a substantial amount of capital has been drawn to the U.S. health care system. The number of nursing home beds increased from about 1.3 million in 1976 to about 1.7 million in 2013. The number of short-term hospital beds (as distinguished from nursing homes) peaked in the late 1970s, at almost 1.5 million, but the total number has since leveled at approximately 915,000 by 2013. There are also considerable and growing amounts of other capital—such as diagnostic equipment—per bed.

Time—The Ultimate Resource

Data on health care expenses and labor and capital inputs reflect only some of the items used by people to produce health. Inputs that are not bought and sold in the marketplace are also important. These include people's own contributions of time and effort in producing health care and entail real costs to society.

For example, when people use their own time to produce better health for themselves, or for loved ones as caregivers, the cost to the individuals and society is the value of the leisure that they forego. Adults who are taking care of their elderly parents for two hours per day, seven days per week, provide care that might otherwise have to be purchased in the market for $15 per hour or more. In this simple illustration, the caregivers provide care worth over $10,500 per year. Though such examples are not necessarily the population norm, these time costs must be added to our measured health care costs.

We have stressed inputs, but the contribution of health resources to the economy is ultimately a measure of the value of the output—health itself. We measure the values of improvements to our health in both consumption and production. We value health both for its own sake and for its contribution to the production of other goods. The intrinsic value of being healthy is ultimately the value we attach to life and limb, which people commonly describe as infinite in certain circumstances, and at least substantial in others. The value of health in the production of other goods is exemplified not just in reduced absenteeism rates but also in output per worker on the job. In both its consumption and production aspects, the output of the health sector makes a substantial contribution to the economy.

The Importance Attached to Economic Problems of Health Care Delivery

The health sector receives attention from policymakers because of its widely perceived problems. The substantial resources devoted to health care are reflected in a more meaningful way through the average level of the nation's spending for health care. Table 1.4 provides various measures of health care spending and its growth since 1960.

Table 1.4 National Health Expenditures and Other Data for Selected Years

Year	*NHE ($billion)*	*% Growth in NHE over Previous Year*	*GDP ($billion)*	*NHE per Capita*	*NHE % GDP*	*CPI*	*Hospital + Related Services Price Index*	*Physician Services Price Index*
1960	27.2		543	146	5.0	29.6		21.9
1970	74.6	13.2%	1,076	355	6.9	38.8		34.5
1980	255.3	15.3%	2,863	1,108	8.9	82.4	69.2	76.5
1990	721.4	11.9%	5,980	2,843	12.1	130.7	178.0	160.8
2000	1,369.7	7.2%	10,285	4,857	13.3	172.2	317.3	244.7
2005	2,024.5	6.7%	13,094	6,856	15.5	195.3	439.9	287.5
2010	2,595.7	4.0%	14,964	8,402	17.3	218.1	621.2	334.1
2011	2,696.6	3.9%	15,518	8,666	17.4	224.9	653.8	343.0
2012	2,799.0	3.8%	16,155	8,927	17.3	229.6	684.0	349.9
2013	2,879.9	2.9%	16,663	9,115	17.3	233.0	701.9	356.5
2014	3,031.3	5.3%	17,393	9,523	17.4	236.7	743.2	361.7

Sources: *NHE and GDP data*: Centers for Medicare and Medicaid Services, NHE Summary Including Share of GDP, CY 1960–2014, www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html, accessed August 2016.

CPI (1960–2014) and price indices (2010–2014): Bureau of Labor Statistics, CPI Detailed Report—June 2016, Tables 24 and 25, www.bls.gov/cpi/cpid1606.pdf, accessed August 2016.

Price indices (1960–2005): U.S. Department of Commerce, Statistical Abstract of the United States, 2012, Table 142—Consumer Price Indexes of Medical Care Prices 1980 to 2010, www2.census.gov/library/publications/2011/compendia/statab/131ed/2012-statab.pdf, accessed August 2016.

Table 1.4 shows how national health expenditures (NHE) grew from $27 billion in 1960 to $3,031 billion in 2014. From 1960, the U.S. population grew from 186 million to 318 million by 2014. Thus, NHE per capita rose from $146 in 1960 to $9,523 in 2014.

However, the real increase is what matters most. Prices, as measured by the broad-based consumer price index (CPI), rose by 700 percent over the same period. After deflating by the CPI, we find that real expenditures per capita in 2014 were 8.16 times the 1960 level—still a hefty increase.[1]

Inflation

Although we have deflated the spending values using the CPI, medical care prices have grown faster historically than prices overall. Table 1.4 also shows the pattern of health care inflation since 1960. Note that hospital and physician care prices have risen much faster than the CPI—a phenomenon that is typical of other health care services and commodities as well.

Medical price inflation is a common problem for maintaining health programs, and it has spurred numerous cost-containment efforts by the government. Understanding and evaluating the effects of such measures are important tasks for the health economist.

Access

For many, the rising costs significantly reduce accessibility to health care. Financial affordability influences demand for most goods and services, and there are many reasons why some people do not have health insurance. What is clear is that the number of uninsured has fallen in response to the Affordable Care Act of 2010. From 2010 to 2016, the uninsured number fell by 20 million people in the United States. The ACA is a form of national health insurance. Later in this book, we will examine several broad groups of plans, the national health insurance programs that exist in other countries, and the newly established ACA.

Quality

Increases in the quality of care contribute to spending increases. Often, the focus is on ensuring quality through professional licensure and certification and, especially for hospitals, through quality-assurance programs. At the same time, concerns arise about access to high-quality care, and they are not limited to those without insurance or with minimal insurance. Other observers, however, express concerns that the quality of care in the United States is often excessive, especially for some "high-tech" treatments. For such treatments, the resource costs may exceed the benefits to patients. The interplay among insurance, technology, and consumption is of major interest to economists.

The Economic Side to Other Health Issues

Production, costs, and insurance naturally involve economics, but many other health issues have economic components, even though they may seem to be purely medical concerns. A few examples illustrate this point.

The choice of a health care treatment seems purely medical to many people, but physicians and other providers increasingly believe in evaluating and comparing alternative treatments on economic grounds. It is necessary to examine the costs of alternative techniques. Physicians are also increasingly sensitive to the economic side of the patient–physician relationship. The patient's preferences are considered valid in determining the appropriateness of a given treatment.

We also must explore the economic reasons behind people's health choices. People take care of themselves well at some times and poorly at other times. People's desired health status can be understood as a meaningful economic choice. Even addiction to a relatively benign substance such as caffeine or a harmful substance such as methamphetamine can be understood better when analyzed as a possibly rational economic choice. Other health issues clearly have an economic aspect: What role should the government play in health? What health care investments should a developing country make? Should cigarette advertising be banned? Questions like these are not solely economic, but they have an economic side.

Economic Methods and Examples of Analysis

We have already provided a formal definition of health economics as "the study of the allocation of resources to and within the health economy." From another perspective, however, health economics is what economists actually do and how they apply economics to health. Economists in practice use certain characteristic approaches to their analyses of the world.

Features of Economic Analysis

Many distinctive features of economics might be exhaustively identified, but we emphasize four:

1 Scarcity of societal resources.
2 Assumption of rational decision making.
3 Concept of marginal analysis.
4 Use of economic models.

Scarcity of Resources

Economic analysis is based on the premise that individuals must give up some of one resource in order to get some of another. At the national level, this means that increasing shares of GDP going to health care ultimately imply decreasing shares available for other uses. The "opportunity cost" of (what we give up to get) health care may be substantial.

While most people will recognize the money costs of goods and services, economists view time as the ultimate scarce resource. Individuals sell their time for wages, and many individuals will refuse overtime work even if offered more than their normal wage rate—because "it's not worth it." Similarly, many will pass up "free" health care because the travel and waiting time costs are too high.

Rational Decision Making

Economists typically approach problems of human economic behavior by assuming that the decision maker is a rational being. We define rationality as "making choices that best further one's own ends given one's resource constraints." Some behaviors may appear irrational. However, when disputes over rationality arise, economists often attempt to point out, perhaps with some delight, that so-called irrational behavior often makes sense when the incentives facing the decision maker are properly understood.

Marginal Analysis

Mainstream economic analyses feature reasoning at the margin. To make an appropriate choice, decision makers must understand the cost as well as the benefit of the next, or marginal, unit. Marginal analysis often entails the mental experiment of trading off the incremental costs against the incremental benefits at the margin.

A prime example involves the purchase of brand-name drugs. Patients' decisions to buy brand-name drugs, particularly for elective treatments, may depend critically on whether they must pay $2 or $3 per pill, or, instead, a fraction of those amounts if prescription drug insurance is available.

Use of Models

Finally, economics characteristically develops models to depict its subject matter. The models may be described in words, graphs, or mathematics. This text features words and graphs. Any model can be pushed too far and must be tested against a sense of reality and ultimately against the facts. Nonetheless, they can be apt, and we can learn from them.

Economic models are often abstract. Abstract models help to make sense of the world, in economics as in everyday life. A young child asking what the solar system is like will

undoubtedly be shown the familiar drawing of the Sun and planets in their orbits—an abstract model. The drawing is quickly grasped, yet no one supposes that the sky really looks like this.

Two Notable Contributors to Health Economics

If health care markets were very similar to other markets there would be no need for a field called health economics. What made this field arise and grow? The defining and distinctive characteristics of the health economy were seen in a seminal work by Kenneth Arrow published in 1963. His exceptional ability in mathematical economics earned him the Nobel Prize in 1972, but his clear thinking about health care markets provided a starting point for health economics. His paper, "Uncertainty and the Welfare Economics of Medical Care," examined how the health care markets differed.

In many endeavors we face win/loss gambles where risks are identified with known probabilities. Arrow pointed out that health outcomes are difficult to predict and may even be difficult to attribute to past behaviors and care. These facts make it complicated to develop markets for risk sharing, and needed insurance markets may fail to develop. To overcome this, partly, health care markets may rely on institutional norms and other institutions such as licensure. The superior knowledge of the physician is relied on as a matter of trust. We must also trust the physician not to base his medical decisions on his own options for profit. The societal norm developed that physicians must remain above such base concerns.

We may say today "So what?" we know all this stuff don't we? Yes, Arrow's insights permeate the thinking of every health economist. But these and his related insights are true, and they make it impossible to see health care markets as just the same as markets for things like "widgets" or ketchup. In consequence this new field has grown and matured.

Amy Finkelstein at MIT has become a preeminent student of health insurance markets, especially Medicare and Medicaid (but also long-term care insurance). Her work created novel theories of choice under risk, but she also conducted large-scale empirical work that fits her theories exceptionally well. An indication of its importance and respect within the wider economics profession is that Finkelstein was awarded the J.B. Clark medal in 2012 for the outstanding economist under the age of 40. Her work has appeared frequently in the best economic journals. She addresses the effect of government insurance programs.

In recent years she has joined with several others to investigate a rich natural experiment in which the State of Oregon gave Medicaid to new recipients on a randomized basis, ideal for scientific research (Finkelstein et al. (2012), Baiker et al. (2013), Taubman et al. (2014)). These data are "rich" because Oregon gave Medicaid access to new recipients on a randomized basis, ideal for scientific research. In short, Medicaid helps recipients to avoid financial disaster. There was also a decline in depression scoring and greater diagnoses of diabetes. More results will become clear as the progress on the study continues.

Does Economics Apply to Health and Health Care?

Many observers complain that economics is irrelevant to the study of health. This issue is raised often enough in serious contexts to require consideration. The complaint suggests a model of health care in which health is primarily a technical issue of medical science, better left to experts. One gets sick and one sees a doctor, who provides the medically appropriate treatment.

If economics studies how scarce resources are used to produce goods and services and then how these goods and services are distributed, then clearly economics applies. Certainly health care resources are scarce; in fact, their cost concerns most people. There is no question that health care is produced and distributed.

Nevertheless, one can question whether the characteristic approaches of economics apply to health care. Are health care consumers rational? Do they calculate optimally at the margin? Imagine a loved one suffering cardiac arrest. Is there time or reasoning power left to calculate? Would anyone question the price of emergency services under such circumstances?

However, much of health care simply does not fit this emergency image. A considerable amount of health care is elective, meaning that patients have and will perceive some choice over whether and when to have the diagnostics or treatment involved. Much health care is even routine, involving problems such as upper respiratory infections, back pain, and diagnostic checkups. The patient often has prior experience with these concerns. Furthermore, even in a real emergency, consumers have agents to make or help make decisions on their behalf. Traditionally physicians have served as agents and more recently, care managers have also entered the process. Thus, rational choices can be made.

An Example: Does Price Matter?

Does price matter? Many have argued that health care is so different from other goods that consumers do not respond to financial incentives. These views have been justified by arguments that demand is based on need, or arguments that patients leave decisions entirely to their providers, who are concerned with their own interests rather than how much patients have to pay.

Data from the RAND Health Insurance Experiment, a pioneering project of the 1970s that examined consumer choices and health outcomes resulting from alternative insurance

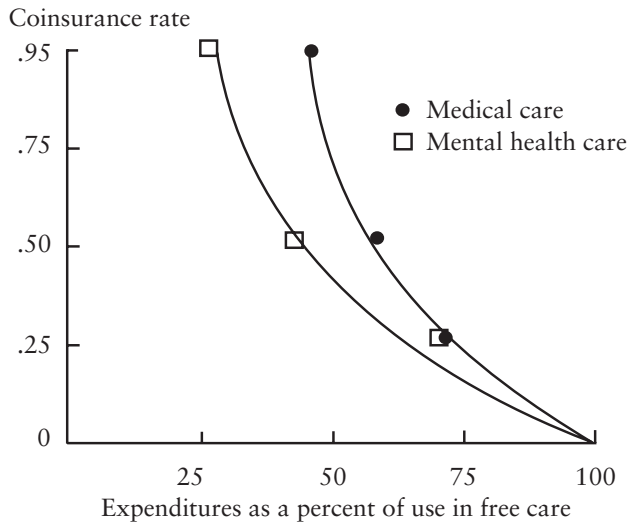


Figure 1.2 Demand Response of Ambulatory Mental and Medical Care in the RAND Health Insurance Experiment

Sources: Keeler, Manning and Wells (1988) for mental health care; Keeler and Rolph (1988) for medical care.

arrangements, give an unequivocal answer to this question: Yes, economic incentives matter. Figure 1.2 examines the use of ambulatory mental health and medical care where amounts of health care consumed are measured along the horizontal axis. These amounts are scaled in percentage terms from zero to 100 percent, where 100 percent reflects the average level of care consumed by the group that used the most care on average. This group, not surprisingly, is the group with "free" care. The vertical axis measures the economic incentives as indicated by the coinsurance rate—the percentage of the bill paid out directly by the consumer. Thus, a higher coinsurance rate reflects a higher price to the consumer.

The curve shown in Figure 1.2 is similar to an economist's demand curve in that it shows people consuming more care as the care becomes less costly in terms of dollars paid out-of-pocket. More importantly, the curve demonstrates that economic incentives do matter. Those facing higher prices demand less care.

Is Health Care Different?

Although economics certainly applies to health care, it is more challenging to answer the question of how directly and simply it applies. Is economic theory so easily applicable that a special field of health economics is not even necessary? Is health care so special as to be unique? Or is the truth somewhere in between?

We argue that health care has many distinctive features, but that it is not unique in any of them. What is unique, perhaps, is the combination of features and even the sheer number of them. We review these distinctive features to alert students as to those salient features of health care that require special attention. In each case where health is distinctive in economic terms, a body of economic theory and empirical work illuminates the issue.

Presence and Extent of Uncertainty

When Arrow directed his attention to the economics of health, he helped establish health economics as a field. He stressed the prevalence of uncertainty in health care, on both the demand side and the supply side. Consumers are uncertain of their health status and need for health care in any coming period. This means that the demand for health care is irregular in nature from the individual's perspective; likewise, the demand facing a health care firm is irregular.

Uncertainty is also prevalent on the supply side. Standard economic analysis often assumes that products, and the pleasures that they bring, are well understood by the purchasers. The purchase of steak, milk, new clothes, or a ticket to a basketball game provides expected well-being that is easily known and understood. In contrast, several cases of product uncertainty exist in the health field. Consumers often do not know the expected outcomes of various treatments without physicians' advice, and in many cases physicians themselves cannot predict the outcomes of treatments with certainty.

Prominence of Insurance

Consumers purchase insurance to guard against this uncertainty and risk. Because we have health insurance, neither most Americans nor citizens of other countries pay directly for the full costs of their health care. Rather, the costs are paid indirectly through coinsurance and through insurance premiums that are often, although not always, purchased through participation in the labor force.

Table 1.5 Personal Health Care Spending, Selected Years (in $ Billions)

Year	*Total Spending*	*Health Insurance*	*Health Insurance*				*Other Third-Party Payers*	*Out-of-Pocket*	*% Out-of-Pocket*
			Private Health Insurance	*Medicare*	*Medicaid*	*Other Health Insurance Programs*			
1960	23.4		5.0				5.5	12.9	55.1
1970	63.1	29.6	14.1	7.3	5.0	3.3	8.5	25.0	39.6
1980	217.0	132.1	61.5	36.3	24.7	9.6	26.7	58.1	26.8
1990	615.3	402.9	204.8	107.3	69.7	21.2	74.5	137.9	22.4
2000	1,162.0	844.2	406.1	216.3	186.9	34.9	118.9	199.0	17.1
2010	2,194.1	1,700.5	754.8	489.8	365.7	90.2	194.1	299.5	13.6
2011	2,280.4	1,772.1	790.6	513.4	373.6	94.6	198.6	309.7	13.6
2012	2,371.8	1,841.6	822.0	534.8	387.8	97.1	211.5	318.7	13.4
2013	2,441.3	1,893.7	834.6	551.2	407.7	100.2	222.1	325.5	13.3
2014	2,563.6	2,000.3	868.8	580.7	444.9	105.9	233.5	329.8	12.9

Notes: Total spending is equal to the sum of health insurance, other third-party payers and out-of-pocket.

Health insurance is equal to the sum of private health insurance, Medicare, Medicaid, and other health insurance programs.

Sources: Centers for Medicare and Medicaid Services, NHE Tables, Table 6—Personal Health Care Expenditures1; Levels, Percent Change, and Percent Distribution, by Source of Funds: Selected Calendar Years 1970–2014, www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccounts Historical.html, accessed August 2016.

Values for 1960 from *Health United States*, 2005, Table 123.

Table 1.5 provides data on the sources of payment for personal health care services for selected years since 1960. In addition to out-of-pocket costs, these payment sources include private insurance; Medicare and Medicaid (the major government programs for the elderly and certain lower income households); and other public and private programs. In 1960, 55 percent of all personal health care expenditures were paid out-of-pocket, meaning that 45 percent was paid by third-party payers (either private or government). Out-of-pocket costs dropped dramatically following the introduction of Medicare and Medicaid in 1966, the continued growth of private insurance, and the introduction of new programs such as the Children's Health Insurance Program (CHIP) established in 1997.

By 2014, 87 percent of personal health care spending was paid by third parties. We will carefully study this phenomenon and its effects for both private and public insurance. It should be clear, even prior to our focused analyses, that the separation of spending from the direct payment for care must weaken some of the price effects that might be expected in standard economic analysis. Insurance changes the demand for care, and it potentially also changes the incentives facing providers.

Changed incentives that face providers concern us more as the insurance portion of the bill increases. How the insurers pay the health care firm thus becomes a critical fact of economic life. Whether insurers cover a procedure, or a professional's services, may determine whether providers use the procedure.

Furthermore, changes in insurance payment procedures can substantially change provider behavior and provider concerns. In the 1980s Medicare, faced with rapidly increasing expenditures, changed its hospital payment system from one based largely on costs (i.e., retrospective reimbursement) to one with fixed payments per admission determined by the resources typically used to treat the medical condition (as classified by Diagnosis Related Groups, or DRGs). With a prospective DRG payment system, an extra day of care suddenly added to the hospital's costs, rather than to its revenues. This reimbursement system, still used today, led to shorter stays, reduced demand for hospital beds, and ultimately the reduction in size and/or closing down of many hospitals.

Problems of Information

Uncertainty can in part be attributed to lack of information. Actual and potential information problems in health care markets raise many economic questions. Sometimes information is unavailable to all parties concerned. For example, neither gynecologists nor their patients may recognize the early stages of cervical cancer without Pap smears. At other times, the information in question is known to some parties but not to all, and then it is the asymmetry of information that is problematic.

The problems of information mean that careful economic analysts must modify their methods. Standard analyses often assume that consumers have the necessary knowledge about the quality of the food or the clothing that they purchase. People purchase beef as opposed to fish, or cotton as opposed to nylon fabrics, basing their decisions on the characteristics of the goods, their prices, and the goods' abilities to bring pleasure.

Health goods and services depart substantively from this model. Consumers may not know which physicians or hospitals are good, capable, or even competent. Consumers may not know whether they themselves are ill or what should be done if they are. This lack of information often makes an individual consumer, sometimes referred to as the *principal*, dependent on the provider, as an *agent*, in a particular way. The provider offers both the information and the service, leading to the possibility of conflicting interests. Newhouse (2002), for example, speaks of a health care "quality chasm" that may be traced to both

inadequate consumer information and to inadequate financial incentives. Health economics must address the provision of health services in this context.

Large Role of Nonprofit Firms

Economists often assume that firms maximize profits. Economic theory provides models that explain how businesses allocate resources in order to maximize profits. Yet many health care providers, including many hospitals, insurers, and nursing homes, have nonprofit status.

What, then, motivates these nonprofit institutions if they cannot enjoy the profits of their endeavors? The economist must analyze the establishment and perpetuation of nonprofit institutions, and understand the differences in their behaviors from for-profit firms. This problem has recently emerged in the context of academic medical centers in the United States. Many current college students, and most certainly their parents and grandparents, know of the prominent roles of great hospitals affiliated with great universities such as Harvard or Johns Hopkins. The public and the larger medical community are aware of the major hospitals as centers of health care, teaching, and research. Yet with the changing health economics of the twenty-first century, the organization of these hospitals and the funding of their activities are continuously evolving.

Restrictions on Competition

Economists and policymakers generally laud the competitive market because the entry of firms or providers in the face of high prices and/or profits will cause the other firms or providers to lower their prices. This entry and the resulting price reduction improve the well-being of consumers.

Nevertheless, the health sector has developed many practices that effectively restrict competition. These practices include licensure requirements for providers, restrictions on provider advertising, and standards of ethical behavior that enjoin providers from competing with each other. We must explain the forces that generated such practices and understand their potential benefits, but we must also understand their anticompetitive impacts and measure the magnitudes of the higher costs they may impose on society.

Regulation to promote quality or to curb costs also reduces the freedom of choice of providers and may influence competition. There is often substantial interest in regulating the health care sector. The causes, as well as the impacts, of the regulations require considerable attention. The pharmaceutical industry, for example, contends that patent protection is crucial for its financial stability. Economists must consider how regulations are developed, as well as who gains and who loses from them.

Role of Equity and Need

Poor health of another human being often evokes a feeling of concern that distinguishes health care from many other goods and services. Many advocates express this feeling by saying that people ought to get the health care they need regardless of whether they can afford it. In practice, "need" is difficult to define, and distributing care under certain definitions of need may cause more economic harm than good. Yet the word signals a set of legitimate concerns for analysis.

Government Subsidies and Public Provision

In most countries, the government plays a major role in the provision or financing of health services. In the United States in 2014, Medicare and Medicaid alone accounted for 40 percent of personal health care spending. However, there are many other government programs, both federal and state and local, including those for public health, military veterans, eligible children, and for mental health and substance abuse. Federal government subsidies are also prominent in the ACA by making insurance coverage more affordable for low and moderate income households.

Conclusions

In this introductory chapter we have sought to explain and support several themes. One is that health and health care markets present a combination of unusual features that together form a unique discipline. Health economists frequently engage in companion disciplines, such as labor economics, public finance, and industrial organization, and recognize that each presents distinctive issues. Health care markets confront risk and uncertainty with unusual information problems. Health professionals have substantial knowledge advantages over their patients. Society norms regarding health and health care make nonprofit motives often preferred, and government provision prominent. Back in the 1960s and 1970s, few economists wrote about health care issues, but now some of the most distinguished economic researchers call themselves "health economists."

Second we have explored the fact that the health economy is "big," so big it is imposing. Nearly 18 percent of GDP in the United States goes to the health sector. Until very recently health care inflation has risen rapidly, raising the question of whether it will begin to gallop again. Prices for health care have grown much more rapidly than consumer prices generally, presenting difficulty for people of modest means to get access. Prior to the ACA, insurance companies often avoided high risk beneficiaries by denying those with pre-existing conditions from buying insurance or by denying them policy renewals.

Finally we have examined the standard methods of economic analysis and suggested how they must be modified to address the characteristics of health and health care markets. While full information is often assumed in introductory microeconomics, it may be asymmetric in health in health insurance markets or even imperfect on both sides. Ordinary firms are often for-profit, while health market norms often prefer nonprofits. A CEO of a business is typically praised for seeking greater profits, while a health professional who does so may not be trusted. And, health economists must study "health bads," where the reduction of the amount consumed is considered an improvement.

The many chapters that follow address these issues. They focus on the ideas, that is the theories, and they describe the empirical work that assesses how well the theories work in reality. These are organized by standard economic categories of demand and supply but also by the relevant health care markets and salient issues of health economics.

The biggest and most important health care issue in America is the Affordable Care Act (ACA), often called "Obamacare." It rivals Medicare and Medicaid in the United States in terms of size and impacts of the reforms. The ACA's principal goal is to reduce the number of people lacking health insurance, which had risen to 50 million people. The ACA also introduced many other reforms that will be described and evaluated in our Health System Reform chapter (22) and elsewhere in the text.

What can economics say about such issues generally? Distinguished economist Victor Fuchs, a past president of the American Economic Association, is optimistic that health economics will meet these challenges and continue to flourish.

> The greatest strengths of economics and economists are a framework of systematic theory, an array of concepts and questions that are particularly relevant to the choices facing policy makers, and skill in drawing inferences from imperfect data. Because health economists often take standard economic theory for granted (like being able to walk or talk), it is easy to underestimate the advantage this framework offers economics over the other social and behavioral sciences. When economists encounter a new problem, one with which they have had no previous experience, they immediately have a way to begin thinking about it long before data collection begins. Scholars in the other "policy sciences" do not. They typically require some detailed knowledge of the particular problem before they can begin to think productively about it. Economists' framework of systematic theory facilitates the transfer of knowledge drawn from other fields of study to the health field.
>
> Health economists have also inherited from economics a set of concepts and questions that have proven to be particularly relevant to the policy problems that have emerged in health during the past three decades. Scarcity, substitution, incentives, marginal analysis, and the like were "just what the doctor ordered," although in many cases the "patient" found the medicine bitter and failed to follow the prescribed advice.
>
> (Fuchs, 2000, p. 148)

Professor Fuchs's insights have become even more relevant following passage of the 2010 Affordable Care Act. These reforms have brought unprecedented change, including an individual mandate for insurance coverage. We share Professor Fuchs's optimism that the theoretical framework and tools used by economists will greatly improve our understanding of these changes and their potential effects.

Postscript

Two significant and related events in November 2016 followed completion of this revised edition.

- First, Donald Trump, who campaigned on the promise to repeal and replace the ACA, was elected President of the United States.
- Second, National Health Expenditure (NHE) data for the United States, published shortly after the election, showed that nominal NHE in 2015 reached $3.2 trillion, an increase of 5.8 percent over 2014. As a share of GDP, NHE increased from 17.4 percent in 2014 to 17.8 percent in 2015.

Our text includes extensive descriptions and analyses of these ACA program provisions and we note that the faster 2014 and 2015 growth rates coincided with major expansions in public and private health insurance coverage under the ACA. While Mr. Trump has pledged to repeal and replace the ACA, many political and health care analysts have speculated

that the major provisions will be difficult to dismantle under a Trump administration. We can only leave it to our students and their instructors to monitor and evaluate any future legislation.

Summary

1 Health care spending has grown rapidly in absolute and relative terms. In 2015, it accounted for nearly 18 percent of U.S. GDP, and its share of GDP is projected to grow.
2 The growth in health care spending is attributable to more services, higher-quality services, and relative increases in the prices of health care services. Health economists seek to determine the underlying causes of these phenomena.
3 The size of the health economy is also reflected through other measures such as the number of jobs in health care professions and amount of capital.
4 Time spent obtaining and providing health care represents a key "unpriced" factor in the health economy.
5 The health economy is considerably larger in the United States, as a share of GDP, than in other countries.
6 There are significant policy concerns not only with the growth of spending but also with access and quality.
7 Economists use models to explain economic behavior. The models are abstract simplifications of reality.
8 Health economists still disagree on some fundamental issues, such as the extent to which the competitive model applies to the health economy.
9 Health care services and the health economy possess a unique set of distinguishing features, such as the prevalence of uncertainty or insurance coverage. Health care is unique because of this entire set of features.
10 The health care system has changed dramatically over the past 50 years. The role of government, reimbursement methods, and the dominance of managed care represent some of the major changes. The Affordable Care Act (ACA) of 2010 is the most important recent change.
11 An important consequence of many of these changes is the substantial drop in out-of-pocket costs for consumers, meaning that private insurance and public programs have correspondingly grown.
12 Technological change through improved procedures, and new drugs, provides potential improvements in health care, but also possibilities of increasing costs, in all countries.
13 Economics provides valuable theoretical tools and a systematic framework for understanding the health care system and evaluating alternative policy proposals.

Discussion Questions

1 Suggest several reasons why health care spending is higher in the United States than in other countries. Is the fact that the U.S. population spends more per capita on health care than people in any other developed country evidence of a failure of the U.S. system? What issues do you think are involved in answering the question?
2 Describe several key issues facing policymakers with regard to health care spending.

3 If greater health care spending leads to more jobs, why is there such concern about the rapid growth rates of spending?
4 Do consumers take the net price (including insurance and time) they face into consideration when choosing health care? What evidence suggests that price matters? Suggest real-life scenarios in which price may affect choices regarding health care.
5 Suppose that a woman works 40 hours per week with no opportunity for overtime. She also takes care of a sick parent. Can we say that her time has no value in providing this health care because she could not earn more at work?
6 What is meant by marginal analysis? Provide an example in which marginal analysis is useful in looking at policy questions.
7 Give three examples of quality of care in the provision of health services. Why might consumers be willing to pay more money to have each of them?
8 Describe the size of the health economy when measured by the quantities of capital and labor used to produce health care. What important inputs to the production of health are not being counted among these?

Exercises

1 Health care spending (S) can be summarized by the following equation:

S = *(population size)* × *(health care quantity per person)* × *(price per unit of health care)*

   (a) Identify three factors that might lead to the rapid growth of health care spending.
   (b) Compare health care spending to housing expenditures and to food expenditures. How are the sectors similar? How do they differ?
2 Identify five distinctive features of the health economy. Examine each one separately, and describe other commodities or sectors that share those features. Do any other commodities or sectors have all the features you listed?
3 In Table 1.1, calculate which countries had the largest and smallest percentage increases in GDP share from 1960 to 1980. Compare these to similar calculations for the period 1980 to 2015. Discuss your results.
4 The United States, Canada, and the United Kingdom share the same language but have considerably different health care systems. Compare the health shares of GDP from 1960 to 2015. Use Table 1.1 to explain the considerable differences among the three countries?
5 Table 1.4 provides indexes of the prices of health care inputs. Calculate the growth rates between 1980 and 2014 of the prices of hospital and physician services. Compare them to the growth rate of the overall consumer price index (CPI). Discuss your findings.
6 In Table 1.5, examine the private health insurance, Medicare, and Medicaid components. Which category grew the most between 1970 and 2014? Between 2000 and 2014? What factors might have led to the differences in the growth rates?
7 Several websites provide useful information on health care and health resources use. Use a Web browser to find sites of:
   - Centers for Medicare & Medicaid Services (CMS)
   - National Institutes of Health (NIH)
   - Organization for Economic Cooperation and Development (OECD)

- The Kaiser Family Foundation (KFF)
- For students outside the United States, find governmental sites from your own country.

Compare and contrast the data available from these sites.

8 The following chart shows health expenditures for the United States between 1960 and 2014. Using a spreadsheet program:

(a) Calculate health expenditures per person for each year.
(b) Calculate percentage increases in health expenditures per person for each year.
(c) Can you find particular events in given years that might explain either small or large changes in the health expenditures per person or in the percentage changes?

Year	*U.S. Population (in Millions)*	*National Health Expenditures ($ in Billions)*
1960	186	27.2
1961	189	29.1
1962	192	31.8
1963	195	34.6
1964	197	38.4
1965	200	41.9
1966	202	46.1
1967	204	51.6
1968	206	58.4
1969	208	65.9
1970	210	74.6
1971	213	82.7
1972	215	92.7
1973	217	102.8
1974	218	116.5
1975	220	133.3
1976	222	152.7
1977	224	173.9
1978	226	195.3
1979	228	221.5
1980	230	255.3
1981	233	296.2
1982	235	334.0
1983	237	367.8
1984	239	405.0
1985	242	442.9

continued

continued

Year	*U.S. Population (in Millions)*	*National Health Expenditures ($ in Billions)*
1986	244	474.7
1987	246	516.5
1988	248	579.3
1989	251	644.8
1990	254	721.4
1991	257	788.1
1992	260	854.1
1993	263	916.6
1994	266	967.2
1995	268	1,021.6
1996	271	1,074.4
1997	274	1,135.5
1998	277	1,202.0
1999	279	1,278.3
2000	282	1,369.7
2001	285	1,486.7
2002	287	1,629.2
2003	290	1,768.2
2004	293	1,896.5
2005	295	2,024.5
2006	298	2,157.0
2007	301	2,296.2
2008	304	2,402.6
2009	306	2,496.4
2010	309	2,595.7
2011	311	2,696.6
2012	314	2,799.0
2013	316	2,879.9
2014	318	3,031.3

Source: Centers for Medicare and Medicaid Services, "NHE Summary Including Share of GDP, CY 1960–2014," www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html, accessed August 2016.

Note

1 The 8.16 multiple is determined by dividing 9,523 (2014 spending) by 236.7 (2014 CPI) and dividing the result by the corresponding ratio for 1960. National health spending updates are available at the Centers for Medicare and Medicaid Services website: www.cms.hhs.gov.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 2

Microeconomic Tools for Health Economics

In this chapter

- Scarcity and the Production Possibilities Frontier
- Practice with Supply and Demand
- Functions and Curves
- Consumer Theory: Ideas behind the Demand Curve
- Individual and Market Demands
- Elasticities
- Production and Market Supply
- The Firm Supply Curve under Perfect Competition
- Monopoly and Other Market Structures
- Conclusions

This chapter provides an explanation of the microeconomic tools used in the text by reviewing material encountered at the introductory and intermediate levels of learning. These tools provide a deeper understanding of ideas in health economics. This review does not substitute for an introductory course in the principles of economics, and a complete series of principles is necessary for a better understanding. Only by such study can one gain an understanding of the subtleties of the subject as well as the many large ideas that we simplified for this review. Although we present the material in a way consistent with more detailed and rigorous treatments in a standard microeconomics course, we have eliminated the academic give-and-take of disputes over issues as well as much of the mathematical detail.

Note also that the "Basic Economics Tools" chapters deal mainly with microeconomic issues; we ignore the macroeconomic issues except in the sense that these subjects increasingly overlap. Microeconomic concerns involve individual decision makers such as households and providers, and specific industries such as the health insurance and hospital industries. Health economics also addresses the problem of the efficient use of scarce resources, which too is a microeconomic issue.

This chapter starts with the concept of scarcity and reviews supply-and-demand analysis. After these introductory treatments of supply and demand, the chapter returns separately to demand first and then to supply, developing the underlying ideas behind these tools. We then combine them into models of market structures, emphasizing those featured in health economics.

Scarcity and the Production Possibilities Frontier

A fundamental idea in economics is that there is no such thing as a free lunch. The fact that little if anything is free implies that to get something, one must usually give up something else, such as time or other resources. A helpful theoretical tool that illustrates this idea is the production possibilities frontier (PPF).

The PPF illustrates the trade-offs between two categories of goods. The curve shows how our choices are constrained by the fact that we cannot have all of everything we want. The idea that we face resource constraints and must make trade-offs is central to the PPF, but similar ideas also apply to the individual firm or the individual consumer.

We begin the production possibilities problem with a table illustrating a classic dilemma concerning society's trade-off between guns and butter. Table 2.1 shows data on the amounts of guns or butter that a hypothetical society could produce with its resources. Guns and butter refer metaphorically to all goods and services with a military use versus those that have a domestic consumption use. The PPF could in principle also be drawn in many dimensions for many goods. What is essential is that the goods represented exhaustively account for all the goods in the economy.

Table 2.1 illustrates two central ideas. Note first that as the number of guns increases, the number for butter falls, indicating that to produce more of one good we must give up some of the other. The amount of butter given up in order to produce an extra 100 units of guns is called the *opportunity cost* of 100 units of guns. We can measure opportunity cost per 100 units of guns as here, but more commonly we look at the opportunity cost of the single next unit of guns, called the marginal unit. In either case, the opportunity cost represents what is given up.

The opportunity cost column reporting the costs of each 100 units of guns in terms of butter foregone illustrates a second idea—that of increasing opportunity costs. As the number of guns increases, the opportunity cost gets larger. If society is to increase its production of

Table 2.1 Society's Trade-Off between Guns and Butter

Point	*Butter*	*Guns*	*Opportunity Cost: Butter Given Up to Produce 100 Units of Guns*
A	936	0	
B	891	100	45
C	828	200	63
D	732	300	96
E	609	400	123
F	444	500	165
G	244	600	200
H	0	700	244

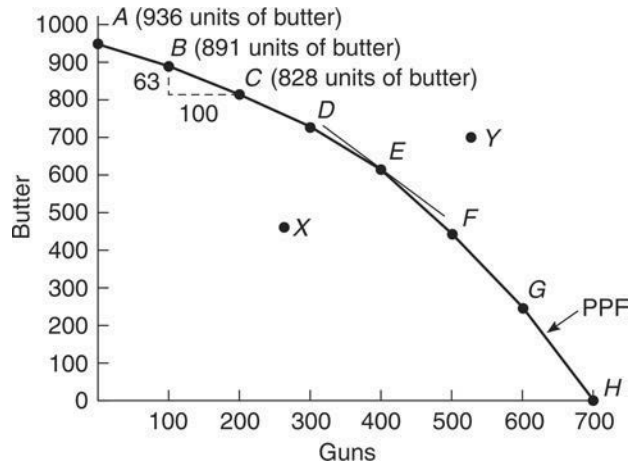


Figure 2.1 Society's Trade-Off between Guns and Butter. Point X Is Inefficient; Point Y Is Infeasible

guns, say from 200 to 300 units, it must transfer the resources, labor, and capital previously used in butter production to gun production. The idea that this is a frontier means that we are representing society's best possible production. Thus, when we first shift butter resources toward gun production, we can arbitrarily choose to shift those resources relatively best suited to gun production first. By choosing laborers who are handier at gun-making than at butter-making, we will gain the most guns per unit of butter we give up. As we shift more resources toward guns, we will have to dig deeper into our relatively good butter-producing resources, and hence give up greater quantities of butter. Increasing opportunity costs also illustrate the specialization of society's resources of labor and capital.

We can transfer the data in Table 2.1 into the graph in Figure 2.1. Note that if this society devotes all its labor and capital to butter production, the most butter it can produce

is 936 units. We represent two numbers—936 units of butter and zero units of guns—by point *A* in Figure 2.1. The other points are transferred in the same manner. We assume that the missing points between these data points will fit the same pattern, resulting in the PPF curve.

This graph illustrates the idea of no free lunch with the downward slope of the PPF curve. In this example, increased gun production means we must give up some butter production—hence, we get no free lunch. Second, the opportunity cost itself is illustrated in the slope of the curve. For example, the line between points *B* and *C* has a slope of 63 (the rise) over 100 (the run), and 63 units of butter per 100 units of guns is the opportunity cost we observed in the table.

The opportunity cost of one single unit of guns is the slope of the PPF at a single point, which equals the slope of a line tangent at that point. Therefore, at point *E* the opportunity cost is identical to the slope of the tangent line to the PPF at *E*. Finally, the idea of increasing opportunity cost is illustrated by the bowed-out shape of the PPF, showing its concavity to the origin. Recall that the slope is the opportunity cost. Thus, the slope becoming steeper means that the opportunity cost is increasing. Society could choose any point on the PPF, but society can be at only one point at a time. How society makes and achieves its choice are other matters to discuss, but at present we have merely illustrated the best possible practices of some hypothetical society. An interior point such as *X* means that the particular society is not doing the best it can; it is inefficient. A point such as *Y* is unattainable because of insufficient resources to produce the indicated amounts of both guns and butter.

BOX 2.1

There's Scarcity and Then There's Real Scarcity

Scarcity is clearly defined in economics; a good or service is "scarce" whenever it has a non-zero opportunity cost. So, goods and services generally are scarce for the very rich, even as much as they are for the very poor.

However, the way ordinary people use the word *scarce* is very clear, too . . . and it can differ from economists' definitions. If money is scarce, that clearly means tough times. Many students may be surprised to discover just how vast the differences among countries are in terms of income, health, and health expenditures. Table 2.2 provides examples from a selection of countries across the globe. Germany and the United States are examples of the industrialized West. In terms of GDP per capita, adjusted for purchasing power parity, these populations experience two to five times the income of the Russian Federation or Brazil, and about four to six times that of either Albania or China. By far the most striking contrast is with the African countries of Ethiopia or Nigeria, who have little to spend on health care, and at the same time experience the largest health and health care deficits.

Table 2.2 There Is a Strikingly Diverse Experience in Income, Health, and Health Care across the World's Cultures

Country	*GDP per capita (in $) 2014*	*Health expenditures per person (in $) 2013*	*Life expectancy at birth (years) total 2013*	*Tuberculosis incidence per 100,000 people 2014*	*Infant mortality rate (deaths per 1,000 live births) 2015*
Afghanistan	634	55	60	189	66
Albania	4,564	240	78	19	13
Bangladesh	1,087	32	71	227	31
Brazil	11,384	1,085	74	44	15
China	7,590	367	75	68	9
Ethiopia	574	25	63	207	41
Germany	47,822	5,006	81	6	3
India	1,582	61	68	167	38
Nigeria	3,203	115	52	322	69
Norway	64,976	6,204	81	8	2
Russian Fed.	12,376	957	71	84	8
United States	54,630	9,146	79	3	6

Source: Table constructed from World Health Organization data archive Global Health Observatory Data Repository. Accessed through http://apps.who.int/gho/data, January 14, 2016.

Practice with Supply and Demand

The most familiar ideas in economics are supply-and-demand curves. We will illustrate supply-and-demand analysis for a hypothetical market for apples.

The Demand Curve and Demand Shifters

In Figure 2.2A, a demand curve illustrates the demand for apples in a hypothetical market for a given period, say one week. The demand curve drawn shows a quantity demanded in this market for this period for each possible price. For example, at a price of $5.25 per bushel, consumers would wish to buy a total of 345 bushels that week. The theory of demand suggests that quantity demanded would be less at higher prices—for example, 215 bushels at a price of $7.50. Some consumers may find that the price rise represents "the last straw" so that they buy none, while others may buy fewer apples than before, and yet others would not change their purchases. It is doubtful that anyone would use the occasion of a price rise as a reason to buy more.

This analysis is done *ceteris paribus*, meaning that we are assuming that all other things are held constant. The price of apples rises while people's tastes, perceptions, incomes, and so on stay the same. In life, it is common for two or more things to change at the same time. If, for example, the price of apples rises at the same time as tastes change, the result would be theoretically ambiguous, meaning that we cannot predict the direction of the change. In contrast, the demand curve depicts the behavior of consumers as price only changes.

As long as people buy less at higher prices, then the demand curve will be downward sloping. In statistical analysis, estimated demand curves are almost always downward sloping. The responsiveness of demand to price is measured by the elasticity. We will discuss elasticity in a later section. Other variables also will affect the demand for apples. For example, the amounts of various foods that people buy may depend on their incomes; richer people tend to buy more. Demand may be affected by the price of other substitutable goods. When any other variable affects demand, its effect will be shown as a shift in the curve. For convenience, we call such variables *demand shifters*. A list of demand shifters includes the following.

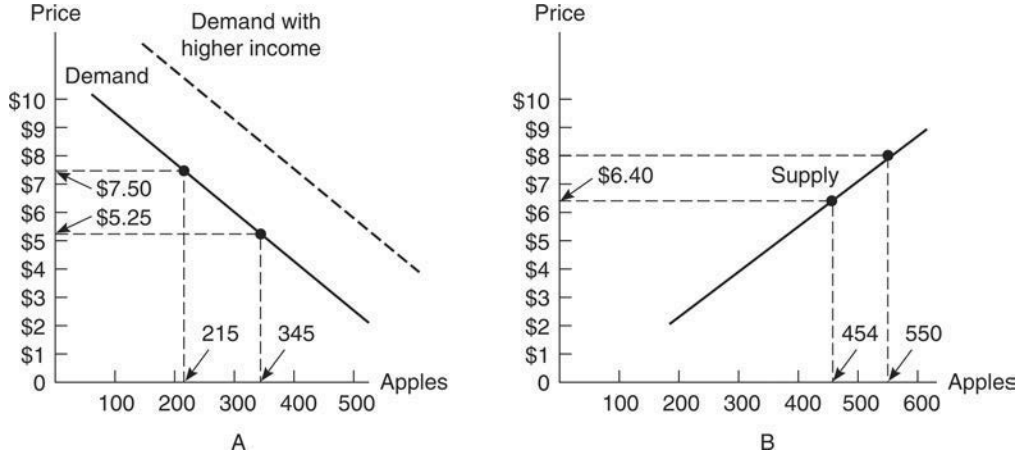


Figure 2.2 Changes in Demand and Supply

Income People with higher incomes tend to demand more of most goods. Such goods are called *normal* goods. But some goods, such as used clothing or generic brand goods, are purchased less often when people become richer. Such goods are called *inferior* goods. Let us assume that apples are normal goods. In Figure 2.2A, increased income in the community would tend to shift the demand curve outward.

Other Prices Prices of related goods also will affect demand. Related goods may be either substitutes or complements. If oranges are regarded as substitutes for apples, an increase in the price of oranges would cause the demand for apples to increase, shifting the demand curve to the right. In contrast, a complement is something that is used with apples, such as caramel. If the price of a complement rises, the demand for apples decreases or shifts left.

Insurance A variable that makes no sense for apples but that is essential on a list of demand shifters in health economics is insurance. Insurance reduces the price to the consumer at the point of service; given the lower price, a greater quantity of health care will be demanded. Although one can treat this as a movement down a given demand curve, we show in a later chapter that this is equivalent to a clockwise rotation in the original demand curve. Insurance plans have many complexities beyond changing the consumer's effective price, and these are also addressed in a later chapter.

Tastes Many other demand shifters may be grouped under the heading of tastes. Tastes can be literally what the word means, as when a new recipe increases interest in apples. The term can be less literal as well, as when we say that an older population has a greater demand for health care because it has a greater taste for health care.

The Supply Curve and Supply Shifters

We approach supply in a similar way. Figure 2.2B shows an upward-sloping supply curve for apples. It illustrates, for example, that apple growers would be willing to offer 454 bushels of apples for sale if the price were $6.40. At a higher price, say $8, more would be offered. Apple growers might be more willing to divert apples from cider production, to make greater efforts in harvest, or even to bring formerly unprofitable trees into production if the price were higher. Such reasons would suggest an upward-sloping supply curve such as the one shown.

We may likewise generate a list of supply shifters.

Technological Change As technology improves for producing a given product, the good becomes cheaper to produce. Certainly, technological changes that make products more costly without improving quality are ignored. As the product becomes cheaper to produce, suppliers are willing to offer more for sale at a given price. This increases supply, thus shifting the supply curve to the right.

Input Prices If the wages of apple pickers were to rise, this increase in an input cost would reduce suppliers' willingness to offer as much for sale at the original price. The supply would decrease, shifting the curve to the left.

Prices of Production-Related Goods The price of a good related in production, such as cider, also would be relevant. Because farmers can use raw apples for eating or for cider, generally a rise in cider prices will cause the supply of apples for eating to decrease, thus shifting the supply curve to the left.

Size of the Industry As more firms (in this case apple growers) enter a market, the supply of the product will be greater. Thus, entry of firms will cause supply to shift to the right.

Weather For a number of products, acts of God such as the weather will tend to affect production. The direction of the effect is obvious: Good weather increases supply.

Equilibrium

Under conditions of competition, the equilibrium in a market will occur at the point where the demand and supply curves intersect. This is the point at which demanders' and suppliers' plans agree. In Figure 2.3, the equilibrium occurs at the price of $5 and at the quantity of 350 bushels. At higher prices, there will be excess supply, and suppliers who were unable to sell all their goods will be willing to bid prices down. At lower prices, there will be excess demand, and demanders who went undersupplied will be willing to pay more and will tend to bid prices up.

Comparative Statics

An equilibrium as depicted in Figure 2.3 is a static equilibrium. It shows a picture of an unchanging equilibrium at a point in time. It is more interesting to assess how the equilibrium will change in response to some economic event. Figures 2.4A and B give two examples. Consider in Figure 2.4A what would happen to the market for coffee if there were a freeze

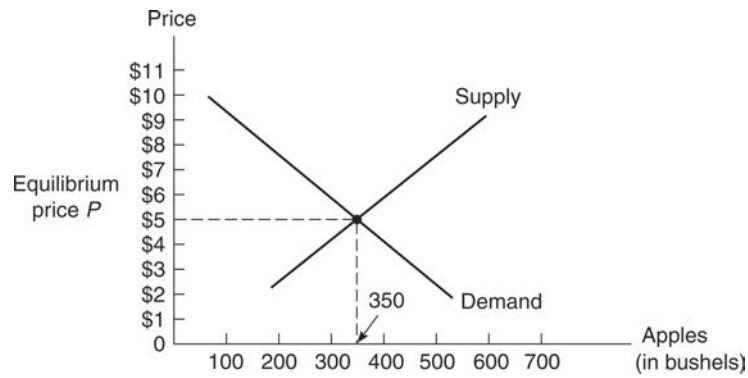


Figure 2.3 Equilibrium Where Demand Equals Supply

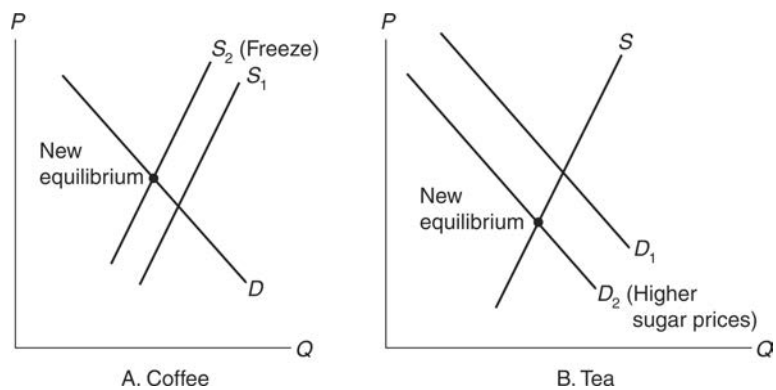


Figure 2.4 Market Effects of Supply-and-Demand Shifts

in Brazil. This worsening of the weather would tend to shift supply to the left as shown. At the new equilibrium, the price of coffee is higher, and the quantity consumed is lower.

Similarly, in Figure 2.4B, consider what happens to the market for tea when the price of sugar rises. Because sugar is a complement, this event causes a shift to the left in the demand for tea as shown. The new equilibrium will have a lower price and a lower quantity.

A few exercises help to generate experience with comparative statistics and to demonstrate the applicability of this analysis:

1 A national health insurance proposal is passed that provides comprehensive health insurance to millions more people than currently. How would this affect the markets for health care in the short run?

   Answer: According to the competitive model, insurance coverage will probably increase on average, causing the demand for health care to increase, shifting the curve to the right. This will increase the equilibrium price of care, as well as the quantity consumed. The result will be an increase in the total money spent on health care. But recall that the analysis is conducted *ceteris paribus*: If an effective cost-control program were put in place at the same time this would reduce the pressure on costs, perhaps cancelling it out.

2 A new law requires that hospitals hire only nurses with baccalaureate degrees. How would this affect the market for hospital care?

   Answer: Hospital markets are not perfectly competitive, but such a law would in effect increase an input price, shifting the supply of hospital care to the left. Under this interpretation, the equilibrium price of hospital care would tend to rise and quantity would fall.

3 Suppose that there is a big breakthrough in the technology for Lasik surgery, that is, surgery designed to correct nearsightedness. Suppose further that this cuts the price of Lasik to a tenth of its previous level with no loss in quality. How would this event affect the market for eyeglasses?

   Answer: Lasik is a substitute for eyeglasses, and demand for eyeglasses would probably decline.

Functions and Curves

Most economic discussions consider the relationships between two or more economic variables. For example, consider what we have theorized about the relationship of the price to the quantity demanded. We say that the quantity demanded is a function of the price. Mathematically, a function is an arrangement whereby we plug in the value of the independent variable, here the price, and the function generates the value of the dependent variable, here the quantity demanded. Alternatively, we can say that quantity demanded depends on price.

Linear Functions

Before considering the writing of supply and demand in functional notation, consider the linear function. A linear function is that of a straight line written as follows:

$$y = a + bx \tag{2.1}$$

where y is the dependent variable and x is the independent variable. A linear function, no matter what idea it represents, has characteristic features: an intercept and a slope, both of

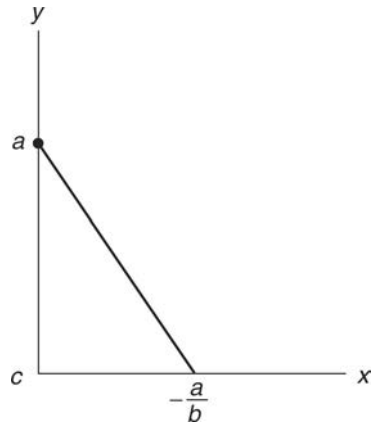


Figure 2.5 Graph of the Function $y = a + bx$

which are constants. The *y*-intercept is the value of the function evaluated when *x* equals zero. Here the intercept is *a*. The slope of a function is the increase in the vertical direction, or the "rise" divided by the increase in the horizontal direction, or the "run."

To determine the slope of this particular function, examine this function drawn as a curve in Figure 2.5. We use the word *curve* for all functions, including the straight line. As noted, the *y*-intercept is *a*. Similarly, the function crosses the *x*-axis at a value of zero for *y*. Setting *y* to zero and solving for *x* yields a value of $x = -a/b$. Now, to find the slope, divide the change in *y*, $-a$, by the change in *x*, $-a/b$, thus generating a value for the slope of *b*, the slope of this linear function. The value of *b* in this case must be negative, as the curve's slope is downward.

Demand Functions

The demand functions up to this point have been linear. In general, though, linear demand is only one special case. Even when demand is linear, there is a minor complication. Consider the following linear demand function:

$$Q_d = a - bP \tag{2.2}$$

where Q_d is quantity demanded and *P* is price.

A complication arises because economists customarily draw demand and supply curves with the independent variable, price, on the *y*-axis and the dependent variable, quantity demanded, on the *x*-axis. In standard mathematics, the dependent variable is usually drawn on the *y*-axis.

Most commonly, we will consider cases where the dependent variable, such as quantity demanded, is a function of not one but several variables. For example, the quantity of spaghetti demanded, Q_d, may depend not only on the price of spaghetti P_s, but also on the price of substitutes for spaghetti (such as other pastas), P_o, the individual's income, *Y*, and a taste factor, *Z*. Mathematically, using the general notation, the demand function for spaghetti would be written as follows:

$$Q_d = f(P_s, P_o, Y, Z) \tag{2.3}$$

Here the notation $Q_d = f(\ldots)$ is read, "Quantity demanded is a function of P_s, P_o, Y, and Z." If the function in (2.3) also happens to be linear, its more specific form would have a characteristic linear look to it. Statisticians frequently use this case, and it is useful to look at an example. A linear spaghetti demand function, for example, might look like this:

$$Q_d = 500 - 10P_s + 5P_o + 20Y + 40Z \tag{2.4}$$

Linear equations with several independent variables have some things in common with the simple linear equation in (2.1). An intercept constant is calculated by setting all the independent variables equal to zero; here the intercept is 500. The slope values in such linear cases will be the coefficients of the independent variables in question. For example, the slope value for the income variable Y is 20. The slope gives information regarding the contributions of changes in the independent variables to the value of the dependent variable. Again, it is worthwhile emphasizing that functions in economics need not be linear. For example, the true spaghetti demand function might instead look like this:

$$Q_d = 7P_s^{20.05}P_o^{0.002}Y^{0.8}Z^{0.01} \tag{2.5}$$

which is not linear. Our theory provides only a few strong conclusions about a demand function: It is downward sloping in its own price, shifting rightward (leftward) with higher prices of substitutes (complements), shifting rightward (leftward) with income increases for normal (inferior) goods, and shifting rightward with a positive shift in tastes. Beyond these features, the demand function mathematically could take on many different forms.

Derived Demand

Demand by consumers for a final good or service may stimulate the provider of that good or service in turn to demand factors of production. There is no theoretical limit to how long this hierarchical chain can be, but the health industry provides several good illustrations. Individual consumers may wish to improve their health or the health of family members. They may demand exercise equipment, healthful foods, and visits to a physician. These consumers generate a derived demand for factors. When considering just a clinic, the manager purchases electronics, technical equipment, drugs, and labor services, as well as the physician's time. The physician and other medically trained personnel, in anticipation of this, had a derived demand for medical education.

Consumer Theory: Ideas behind the Demand Curve

Consumer theory examines how rational individuals make consumption choices when faced with limited resources. The limited resources determine what options a consumer can afford. From among these options, the consumer attempts to pick the best one. The theory has two parts. One is a description of what the consumer prefers—what he or she thinks is best; for this description, we use the ideas of utility and of indifference curves. The second part is a description of what the consumer can afford; for this part, we use the idea of budget constraints. The use of indifference curves and budget constraints together constitutes indifference curve analysis.

Utility

Consider Kathy Richards, an experienced consumer who knows what her tastes are for kinds of houses to live in, cars to drive, food to eat, and books to read. She can't afford everything she would like to have, but she knows what she would prefer if she could afford everything.

In summarizing this information about Kathy's preferences, we suppose that she has a utility function. Utility is a measure of her level of satisfaction with various combinations of consumer goods. It includes a market basket filled with a combination of housing, food, transportation, and so on, with perhaps many types of each. We assign a greater value of utility to bundles preferred over other bundles. Because more utility, thus defined, is always better, Kathy will logically seek to maximize her utility subject to the constraint of what is affordable to her.

Using functions, we say that Kathy's utility is a function of the goods and services she consumes. In practice, the level of detail we use will vary. On some occasions, we must specify most of the detailed consumption of Kathy's life. Then, we would describe her utility as a function of each good or service she buys, perhaps compiling hundreds of them. But in many cases, it is useful to abstract from this detail and describe Kathy's utility as a function of one or two goods of special interest, plus another general good representing all other goods she considers. At still other times, we will find it useful to describe Kathy's utility as a function of wealth.

We will develop two examples in the following discussion. Theories using the idea of utility may propose that utility is either cardinal or ordinal. Cardinal utility means a metric measure, like a measure of weight or volume. It is characteristic of cardinal measures that the difference as well as the ratio between two measurements has meaning. One can meaningfully say, for example, that five quarts is three more than two quarts, and, for that matter, it is two and one-half times two quarts. Under ordinal utility, to the contrary, only the ranking has meaning. Examples of ordinal numbers are first, second, third, and so on.

It is generally preferable to theorize that consumers' utility is ordinal. It is safer to assume that consumers can rank their preferences than to assume that they can both rank and scale them. It seems safer to suppose someone can say that he or she is happy to have gotten a raise, but it seems questionable to suppose that he or she is 1.07 times as happy. Most theories of demand assume only ordinality. In a few cases, such as the theory of behavior under risk and insurance, some degree of cardinality is assumed.

To illustrate ideas about utility, we should begin with the simplest case to draw. Figure 2.6 depicts Kathy's utility as a function of her wealth. The curve illustrates two ideas. First, the upward slope indicates that utility increases with wealth. Second, the bowed shape, concave from below, indicates that her utility increases with wealth at a decreasing rate.

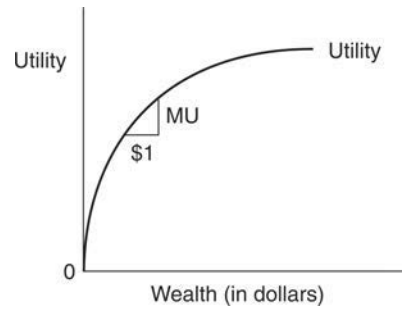


Figure 2.6 The Utility of Wealth

Marginal utility is defined as the extra utility achieved by consuming one more unit of a good. Here, the only good is wealth so marginal utility is the extra utility Kathy gets from one more dollar of wealth. An extra convenience of drawing the function graphically is that the marginal utility is the slope of the curve at a given point. Starting at any point on the curve and adding a single dollar to Kathy's wealth leads to extra utility for that dollar, which we have just defined as marginal utility (MU).

If we understand marginal utility as the slope, the marginal utility of wealth for Kathy gets smaller as she gets wealthier. That is, the slope gets flatter. An extra dollar means more to Kathy when she is poorer than when she is richer. Does this notion apply to most people? That it might apply seems plausible to most students, but the notion also introduces an element of cardinal utility. Cardinal utility is essential when analyzing consumer decisions regarding risk and insurance, discussed in detail in Chapter 8.

Indifference Curves

Often, we wish to depict the consumer's preferences over two or more goods. The most convenient case to draw is when there are only two goods. To capture the sense of the real world in a two-good drawing, let one of the two goods represent all other goods generally, as if they were a conglomerate. Call this conglomerate OG, meaning Other Goods. Suppose that the good of special interest is Food. Figure 2.7 depicts a graph with these two goods, OG and Food, on the axes. Any point in the space, such as *A*, represents a consumer bundle. The bundle *A* includes the combination of eight units of Food and four units of Other Goods. Other bundles that are labeled include *B*, *C*, *D*, *E*, *F*, and *G*, but any other point in the space is also a bundle.

Suppose we focus on bundle *A* and that we hypothetically ask Kathy to identify all other bundles as well that for her are indifferent to *A* (that is, points that give her the same utility as *A*). The entire set of such points is labeled U_1; as noted in the graph, any point along U_1 affords Kathy 112 utils (we will call the units of utility *utils*).

This curve is downward sloping, as well as bowed toward the origin. Notice, for example, that Kathy did not choose point *C* as being indifferent to *A*. This seems plausible because *C* represents more of both goods, and as long as she is not satiated with these two goods, then she would prefer *C* to *A*. Likewise, she has not picked point *D* as indifferent to *A* because

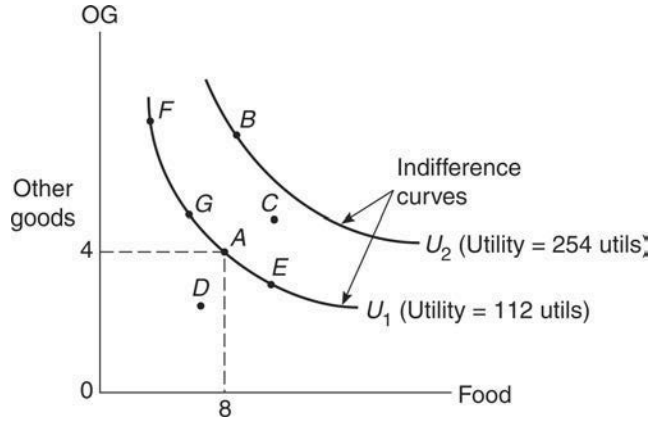


Figure 2.7 Indifference Curves between Food and Other Goods (OG)

D has less of both goods. Instead, she has picked points such as *E*, which has more Food but less OG. Presumably, at *E* she has just balanced the loss of OG against the gain in Food. These considerations suggest that it is understandable that the indifference curve through *A* is downward sloping.

The curve is also convex to the origin. Consider that at point *F*, Kathy has relatively a lot of OG and little Food. As Food is relatively scarce for her, she is willing to give up a lot of OG to get more Food. We describe the rate at which she is willing to trade off the two goods by the slope of the indifference curve, which is steep at point *F*. In contrast, as we move down the indifference curve, Kathy gains relatively more Food, and the more she gets, the less ready she is to give up still further OG to gain yet more Food. Thus, the curve becomes flatter.

Indifference curves for Kathy summarize and represent her preferences. Every possible combination of goods will lie on some indifference curve so that in principle there would be an infinite number of indifference curves in Figure 2.7, with higher curves representing greater satisfaction.

Budget Constraints

Indifference curve analysis uses preference maps and budget constraints. The budget constraint indicates the set of bundles affordable with a given income. Suppose that Kathy must allocate $30 of her family food budget per week between beef *B* and chicken *C*. If the price of beef is $2 per pound, and the price of chicken is $1 per pound, then she can afford any combination of *B* and *C* that costs less than or equal to $30, and her budget constraint is:

$$30 = 2B + 1C \tag{2.6}$$

If we draw this constraint in Figure 2.8 with beef *B* on the vertical axis, then the budget constraint will start at 15 pounds of beef and proceed downward sloping to 30 pounds of chicken as the horizontal intercept. It is convenient to demonstrate this pattern by examining the mathematical function. Equation (2.6) can be transformed using algebra so that *B* appears on the left-hand side and all other terms are on the right. Thus, the equivalent budget constraint is:

$$B = \frac{30}{2} - \frac{1}{2}C = 15 - 0.5C \tag{2.7}$$

a linear function with an intercept of 15 and a slope of –0.5. If, in Figure 2.8, the price of chicken rose, the amount that Kathy could buy, if she spent all $30 on chicken, would be less than before. If the price doubled, the chicken axis intercept would shift inward, permitting her to buy only 15 pounds of chicken rather than 30. The beef intercept is not affected when the price of chicken rises.

Consider instead an increase in the portion of her budget allocated to beef and chicken. A doubling to $60 would allow Kathy to increase the amount of beef from 15 to 30 pounds, or the amount of chicken from 30 to 60 pounds. As shown, the new budget constraint lies parallel to the original budget constraint. Doubling the income in itself does not cause the prices to change. Because the slope of each budget constraint is the ratio of prices, the new constraint will be parallel. The intercepts will double.

Consumer Equilibrium

To maximize satisfaction given a budget constraint, the consumer will seek the highest attainable indifference curve. In Figure 2.9, the indifference curve U_1 is not the best possible, while

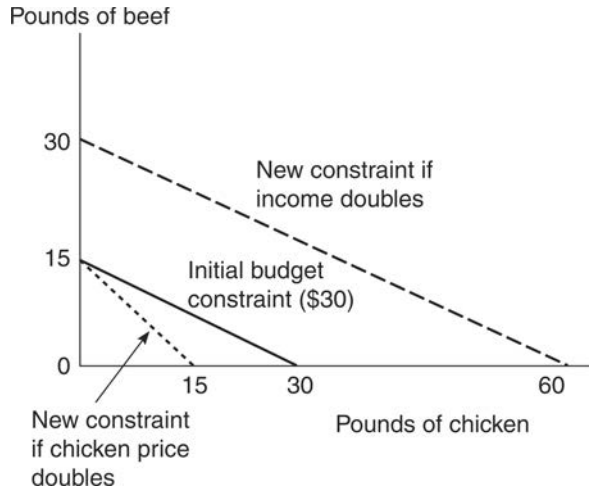


Figure 2.8 Changes in Budget Constraints Due to Changes in Price or Income (OG)

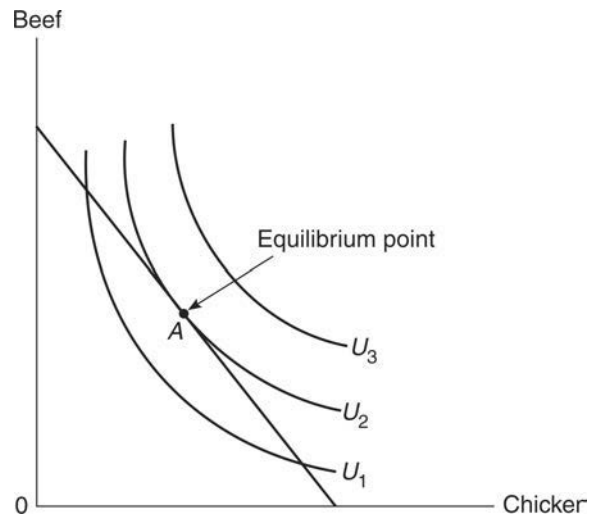


Figure 2.9 The Consumer's Equilibrium

the indifference curve U_3 is unattainable. Rejecting such alternatives, the consumer will find that she maximizes utility at a point of tangency, shown as point *A* in the figure. At this point, the rate at which the consumer is willing to trade beef for chicken, the slope of the indifference curve equals the rate at which the consumer is able to trade the two goods at market prices—the slope of the budget constraint.

One can derive the equilibrium for different prices and/or for different values of income. For example, in Figure 2.10A, as the price of chicken rises consecutively, Kathy consecutively chooses points *A*, *B*, *C*, and *D*. The collection of such points is called the *price offer curve*.

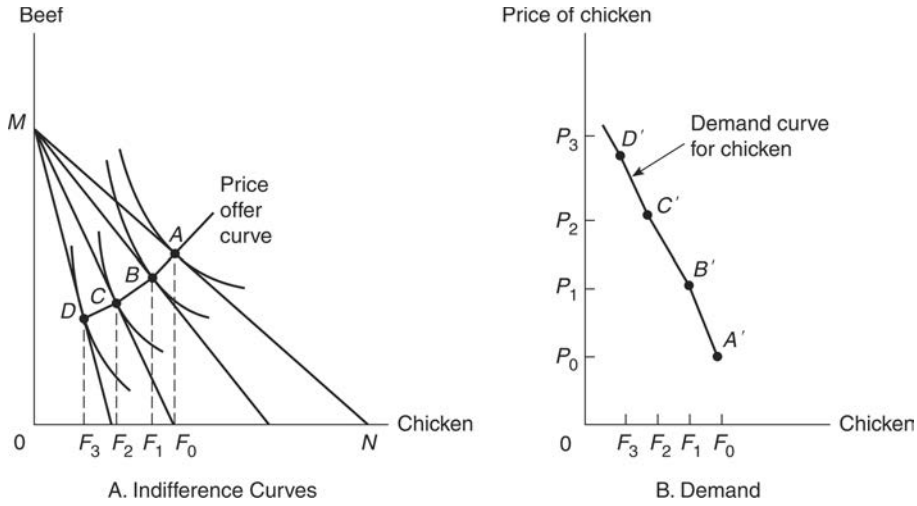


Figure 2.10 Derivation of the Consumer's Demand Curve

The information from these data points along the price offer curve can be used to generate her demand curve. For example, point *A* in Figure 2.10A represents her best choice for a given budget and set of prices. Suppose the price of chicken for budget constraint *MN* is P_0. Then, plotting P_0 together with the quantity demanded at *A*, namely F_0, in the graph at right generates point *A'*. In the same manner, data on price P_1 and the quantity demanded F_1, at point *B*, generate point *B'* in the graph at right. Repeating this process generates the demand curve.

Individual and Market Demands

The theory of consumer behavior focuses on the demand relationships of individual consumers. Many applications of demand theory, however, consider market demand. The extension from individual to market demand is straightforward. In Figure 2.11, health care demand is shown for two people who constitute a market for health care. For example, Mary demands 3 units at the price $p_H = 30$, and John demands 2 units. Market demand here at $p_H = 30$ is (3 + 2) or 5 units. The market demand in panel C is derived by adding the quantities demanded at every price. The process can be extended to all of the consumers in the market, and it yields a market demand curve. If everyone's demand curve conforms to the law of demand, the market demand curve must also be downward sloping.

Finally, note that as with individual demand functions, other variables such as income and the prices of related goods (the shift variable we discussed earlier) affect market demand. Thus, the market demand for some commodity *X* might be expressed in functional notation. Consider, for example

$$QD_x = f(P_x, Y, P_o, E) \tag{2.8}$$

where *Y* represents income, P_o represents the prices of other goods, and *E* represents a socioeconomic variable such as average educational attainment (in years of schooling).

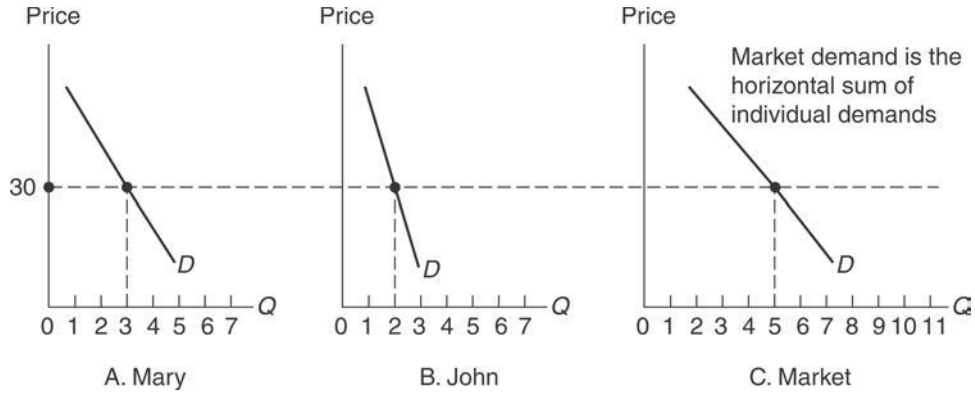


Figure 2.11 Derivation of a Market Demand Curve

Elasticities

We often seek to understand the responsiveness of the quantity demanded to other variables. For example, if the price of health care rises, will the quantity demanded fall by a large amount or a small amount? Economists use the term *elasticity* to describe the responsiveness of any term y (in this case, quantity demanded of health care) to changes in some other variable x (here, price of health care).

Elasticity is defined as the percent change in the dependent variable resulting from a one percent change in the independent variable under study. Percentages allow us to "standardize" our measure and to eliminate problems comparing different goods or different units of measurement.[1] In the case of the price elasticity of demand, it is as follows:

$$E_p = (\%\text{ change in quantity demanded}) \div (\%\text{ change in price})$$

or

$$E_p = \left(\frac{\Delta Q / Q}{\Delta P / P}\right) = \frac{\Delta Q}{\Delta P}\left(\frac{P}{Q}\right) \tag{2.9}$$

where Δ refers to change in the variable. The price elasticity is always algebraically negative because an increase in price leads to a decrease in quantity demanded. We derive other elasticities, such as the income elasticity of demand, similarly.

$$E_y = (\%\text{ change in quantity demanded}) \div (\%\text{ change in income})$$

or

$$E_y = \left(\frac{\Delta Q / Q}{\Delta Y / Y}\right) = \frac{\Delta Q}{\Delta Y}\left(\frac{Y}{Q}\right) \tag{2.10}$$

Income elasticity may be positive (if a normal good) or negative (if an inferior good). If a variable elicits no response at all, then elasticity is zero.

Numerical values for price elasticities are often reported in absolute values, eliminating the minus sign. Absolute values for price elasticities indicate the responsiveness of demand to price in that the greater the elasticity, the greater the responsiveness. Absolute values greater than 1 are considered relatively responsive and are called elastic. Elasticities less than 1 in absolute value are called inelastic. In the limiting cases, zero price elasticity means that the demand curve is perfectly vertical, while infinite price elasticity means that the demand curve is perfectly flat.

The importance of price elasticity to policy questions can be illustrated with an example regarding cigarettes, which are clearly a health concern. Suppose that a state added 50 cents per pack to its tax on cigarettes. Together with supply-and-demand analysis, our elasticity concepts help us identify the main policy issues. Lawmakers hope that such a tax increase will curb smoking and bring in tax revenue, but these tend to be contradictory goals. The exact effects will be difficult to predict unless reliable estimates are available of the cigarette price elasticity. If one discovered that demand is perfectly inelastic (D_1 in Figure 2.12), tax revenue would be at a maximum but with no effect on smoking or health. Alternative scenarios of increasingly elastic demand (D_2 and D_3) create bigger reductions in smoking but at the cost of decreasing tax revenues. Thus, the more elastic the response, the greater the effectiveness of an excise tax in inducing people to reduce their levels of smoking. Lewit and Coate (1982) indicate that teenagers, for example, are more responsive to cigarette prices than are adults. In such cases, taxes on cigarettes will be relatively more effective with teenagers.

Market demand elasticities vary by industry and by product. Those goods and services that we call necessities tend to have elasticities less than 1 in absolute value, while luxuries are more elastic. Short-run elasticities are generally smaller in absolute value than long-run elasticities. Further, goods that cost only a tiny fraction of one's income motivate little or no

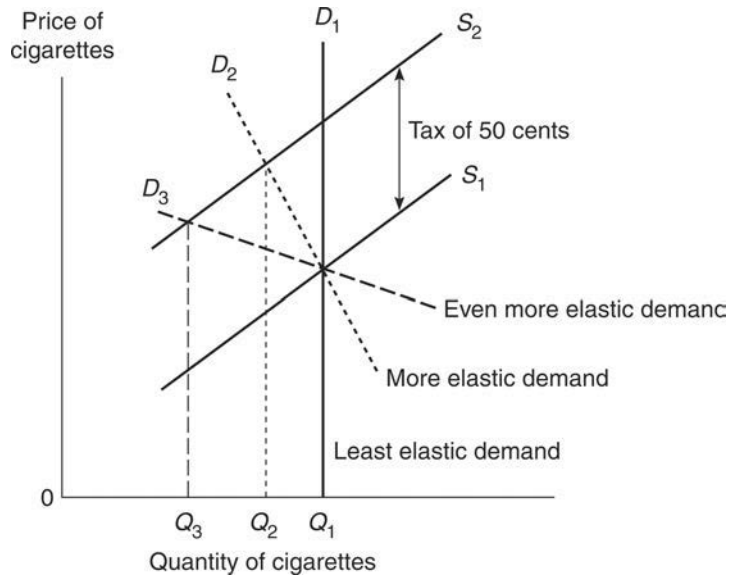


Figure 2.12 The Impact of a Cigarette Excise Tax

Table 2.3 Demand Price Elasticities for a Variety of Goods

Good or Service	*Price Elasticity*
Hospital Care[1]	–0.14 to –0.17
Physician Care[2]	–0.16 to –0.35
Apples (U.S.)[3]	–1.15
Bread (U.K.)[3]	–0.26
Gas, Short Run (Canada)[3]	–0.01 to –0.20
Cigarettes (U.S.)[4]	–0.30 to –0.50
Beer (U.S.)[5]	–0.20 to –0.40
Beef (U.K.)[3]	–1.45
Motion Picture Tickets[6]	–3.40
Marijuana[7]	–0.67 to –0.79

Sources: [1] Wedig (1988); [2] Manning (1987); [3] Mansfield et al. p. 103 (2002); [4] Keeler (1993); [5] Grossman (1998); [6] Ruffin and Gregory, p. 102 (1997); [7] Davis, Geisler, and Nichols (2015).

"shopping around," making their demand elasticities very small in absolute value. Table 2.3 provides some common estimates of demand elasticities for a variety of products, and readers can compare these findings to their own personal experiences.

Production and Market Supply

If market demand is one "blade of the scissors" in determining the price of a good, market supply by individual firms is the other. A typical producer, or firm, faces decisions on production levels, prices, production methods to use, levels of advertising, and amounts of inputs to purchase. The theory of the firm, much like the theory of consumer behavior for buyers, develops a framework for understanding these choices.

The key assumption for most models of firm behavior is that the decision makers wish to maximize profits. It follows that the firm will try to minimize the costs of producing any given output and will undertake activities, such as advertising, only if they add to profits. Before examining such decisions, we will review production and cost relationships.

The Production Function

The production function shows the maximum sustainable output that can be obtained from all of the various possible combinations of inputs such as labor, materials, and machinery, with existing technology and know-how. We begin our discussion of production functions with a simple case—one in which there is just one input and one output. Suppose that food, perhaps in a hunter-gatherer society, was produced solely with labor. We show the production function in the top panel of Figure 2.13. The fact that the production function is rising indicates the idea that labor is productive; more labor means that more food is produced. The bowed shape of the curve illustrates a second idea—the law of diminishing returns.

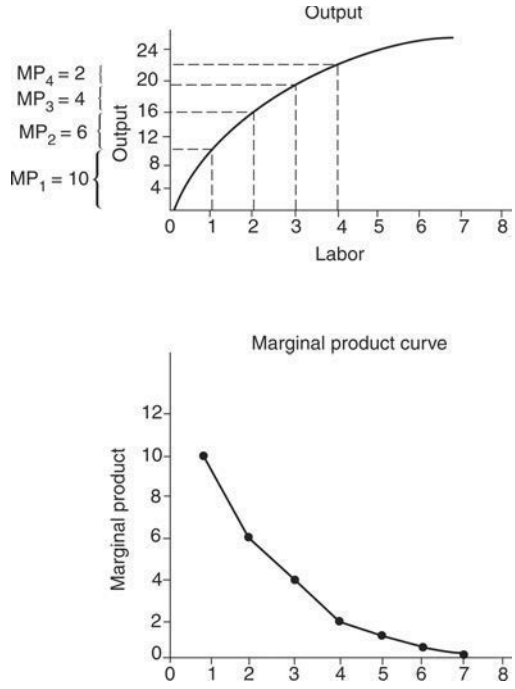


Figure 2.13 Output and Marginal Product of Labor

Students should understand that in Figure 2.13 production takes place during a specified period of time. Thus, the output axis represents a flow of output per unit of time. Likewise, the labor input represents labor services applied during a specified period of time. The law of diminishing returns represents the idea that the marginal product (MP) of an input will eventually tend to fall as more is added. The MP (in this example, labor) is defined as the extra output that can be generated when one adds an additional hour of labor, holding all other inputs constant. In the figure, the food output increases from zero to 10 units when the first hour of labor is added. Thus, 10 units of food is the marginal product of the first unit of labor. When a second hour of labor is added, the output of food increases from 10 units to 16 units. The extra amount is six units, meaning that the marginal product of the second hour of labor is six units of food. The bottom panel of the graph illustrates the marginal product (MP) of consecutive hours of labor. The pattern of the MPs in this illustration is clear: They tend to get smaller as more labor is applied. This illustrates the law of diminishing returns. Notice that total output need never fall during diminishing returns. That is, the production curve itself never turns downward in this illustration, although it may in some applications.

Production Functions

In practice, production processes may involve several inputs, not just labor. It is convenient to express the production relationship for a firm, or a unit of the firm, as follows:

$$Q = f(X_1, X_2, \ldots, X_n) \tag{2.11}$$

Here Q represents output; X_1, X_2, and so on are quantities of the various inputs. The ultimate output of the health industry is health, and its related production issues are termed the "production of health," a phrase that applies not only to larger populations but also to the consumer's production of his or her individual health. When individuals seek to improve their family's health, they often purchase health care from hospitals, clinics, nursing homes, home health agencies, and so on. One step removed logically from health production, these elements of health care must themselves be produced, and much economic research and concern are directed at this production stage, too.

Consider, for example, the output of hospital X-ray services, which require labor by technicians, nurses, and radiologists, and machinery such as X-ray machines, computers, and film. As with demand functions, production functions may take on many mathematical forms. The theory of production specifies only certain patterns for these functions.

One commonly applied functional form that fits the theoretical patterns for such functions is the Cobb-Douglas form. Historically, it was one of the earliest production functional forms to be studied and applied to firms. It derives its name from mathematician Charles Cobb and economist (and later U.S. Senator) Paul Douglas. Many other functional forms of production have since been investigated, but this form is still commonly used in the classroom to illustrate the mathematics of the production process.

If the production of X-ray services just discussed fits the Cobb-Douglas form, and if inputs of all kinds are grouped into the categories of capital, K, and labor, L, the production function actually estimated might look like this:

$$Q = L^{0.8} K^{0.2} \tag{2.12}$$

The exponents have natural interpretations. A 10 percent increase in labor increases quantity by eight percent (0.8×10); a 10 percent increase in capital has a two percent impact on quantity. In this particular case, 10 percent increases in both capital and labor thus increase quantity by 10 percent.

Here, as with any production function, there will be a unique level of maximum output for any combination of inputs. For example, Table 2.4 shows values of output corresponding to changes in L, holding K constant at five units. The change in output associated with a one-unit change in L is the marginal product of labor. More L (e.g., technicians) with the same amount of K will typically produce more Q (hence, a positive marginal product of labor), but as the machines become crowded or break down, the marginal product may fall as L increases. Hence, the marginal product in Table 2.4 diminishes. This decreasing marginal product again illustrates the law of diminishing returns. The average output or average product (AP) for each worker is Q/L, as shown in the last column.

Alternatively, we can derive various input combinations needed to produce a given output level. Table 2.5 illustrates several combinations that produce 10 units of output for the production function represented by equation (2.12). This method closely parallels the indifference curve analyses introduced in an earlier section, except that we actually observe and measure the quantities produced (in contrast to the levels of utility that could only be ranked). Combinations of inputs producing equal output levels lie on an isoquant (literally, "the same quantity"). The isoquant in Table 2.5 is illustrated graphically in Figure 2.14. The isoquant map, representing all possible values of Q, would be the geometric representation of a production function.

Table 2.4 Production Schedule for X-ray Services

K	*L*	*Q*	*MP*	*AP*
5	0	0.00	—	—
5	1	1.38	1.38	1.38
5	2	2.40	1.02	1.20
5	3	3.32	0.92	1.11
5	4	4.18	0.86	1.05
5	5	5.00	0.82	1.00
5	6	5.79	0.79	0.97
5	7	6.54	0.75	0.93

Table 2.5 An Isoquant Schedule

Q	*L*	*K*
10	1	100,000.00
10	5	160.00
10	7	41.60
10	8	24.40
10	10	10.00
10	11	6.80
10	13	3.50
10	20	0.63

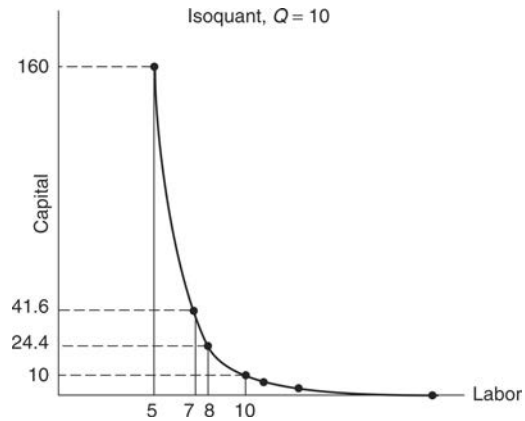


Figure 2.14 A Production Isoquant

The negative slope to an isoquant indicates the possibility of substituting inputs in the production process and of the positive marginal product of the inputs. Consider, again, the example of X-ray services. The numerical value of the isoquant slope, indicating how much capital (X-ray machines, film, computers) must be given up, per unit increase in labor (nurses, technicians, radiologists), is called the marginal rate of technical substitution of labor for capital ($MRTS_{LK}$).

IS SUBSTITUTION POSSIBLE IN PRACTICE? Empirical estimates reveal substantial substitution possibilities between physicians' assistants and physicians. Other studies reveal substitution between nurses and residents and between hospital capital and hospital staff. How can physical capital substitute for a human medical practitioner? Later chapters discuss these issues.

Isocost Curves

In order to maximize profits, those running the unit will want to minimize the cost of producing any given output. Letting TC represent total costs, w the price (wages, salaries, fringe benefits) of labor, and r the cost of buying or renting machines for the production period, the total cost is as follows:

$$TC = wL + rK \tag{2.13}$$

where L and K are the amounts of inputs used, labor and capital. For example, if $w = 50$, and $r = 20$, when the unit uses 30 hours of labor and 10 machines, $TC = 1{,}700 = (50 \times 30) + (20 \times 10)$. As with the consumer's budget problem, it is helpful to determine all of the combinations of L and K that cost a given amount, such as \$1,000. The equation for this isocost curve is

$$1{,}000 = 50L + 20K \tag{2.14}$$

Again, as with the consumer example, we can rearrange the equation by placing capital (or labor) on the left-hand side, to yield

$$K = 50 - 2.5L \tag{2.15}$$

which is a linear equation as shown in Figure 2.15. The isocost curve for $TC = 686$ is also shown. More generally, equation (2.15) can be written as

$$K = TC/r - (w/r)L \tag{2.16}$$

Equation (2.16) shows the impacts of changes in wages and/or rental rates on the costs of purchasing various amounts of labor and/or machinery.

Cost Minimization or Output Maximization

In the example just given, the assumption that firms maximize profits requires that the X-ray unit select the least-cost method of producing its output. To produce Q^* units of output, the firm will wish to minimize the costs of that Q^* output, and hence be on the lowest possible isocost curve. The case where Q^* equals 10 units is illustrated in Figure 2.15. This will occur at point A, where the isocost curve is tangent to the isoquant representing 10 units of output. Higher isocost curves are unnecessarily wasteful; lower ones will not attain 10 units of output.

Suppose instead that the firm has a budget of exactly \$686 and wishes to maximize its output. It is easy to see from Figure 2.15 that the same equilibrium condition will hold, and

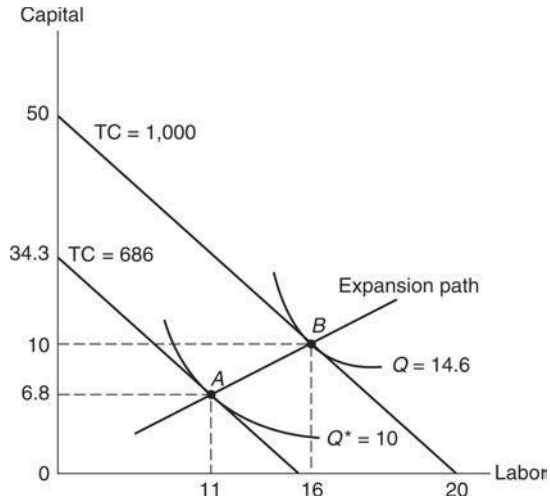


Figure 2.15 Cost Minimization (Output Maximization) Determining Efficient Combinations of Labor and Capital

that the most the firm can produce is 10 units. Cost minimization and output maximization in the manner described lead to the same results.

LOOKING AT COSTS FROM A DIFFERENT VIEWPOINT Cost minimization can also be looked at from society's point of view. For example, hospitals may achieve cost minimization in applying hospital inputs, and home health care services may achieve cost minimization in applying home health care resources. But, for society as a whole to minimize its costs of care, we need to know which of these types of care, home health or inpatient hospital care, is the most cost efficient for particular patients and we need to know if the quality of care is comparable.

Marginal and Average Cost Curves

By varying the production levels and finding the respective isocost curves, we can find the minimum cost of producing each output level. This is shown by the set of tangencies in Figure 2.15. The curve connecting these tangency points, *A* and *B*, is called the expansion path. Thus, the expansion path contains the information on the total cost and the average cost (cost/unit) of producing any output level.

If all inputs can be varied, then the long-run total cost and long-run average cost (LRAC) functions are generated. The LRAC curve is illustrated first in Figure 2.16. Total and average costs are related to the scale of the activity. If higher levels of production lead to improved ability to take advantage of specialization providing a better division of labor, it may be possible to reduce average costs; the case of decreasing long-run average costs is referred to as the case of economies of scale. If, on the other hand, the increased level of output leads to difficulties in managing and coordinating production activities, then long-run average costs may rise; this is referred to as the case of diseconomies of scale. Such issues are relevant for determining the optimal size for firms. For example, the socially optimal size and distribution of hospitals will depend on estimates of scale economies. As another example, it is clear that

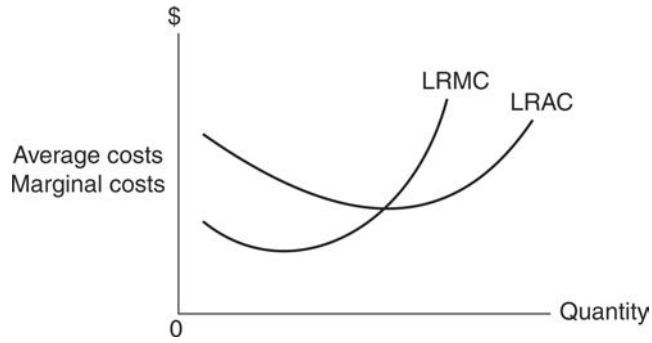


Figure 2.16 Long-Run Cost Curves for a Firm

enough patient volume is needed to cover costs of such high-priced items as CAT scanners, a case of economies of scale. However, too many patients may lead to crowding of patients or to increased labor costs that could again increase costs, producing a case of diseconomies of scale.

In our specific Cobb-Douglas production function example, the LRAC will be a horizontal line reflecting constant average costs (about $68.60 per unit). This occurs simply because of the specific production function chosen as an example. Figure 2.16 shows the classical U-shaped relationship, which starts with economies of scale and then yields to diseconomies of scale. The long-run marginal cost (LRMC) curve shows the cost of producing an incremental unit when all inputs (both machinery and labor) can be varied. It will go through the minimum point of the LRAC.

The short run corresponds to a period where at least one input (typically machinery or plant) cannot be changed. This analysis applies particularly to big-ticket machinery items in hospitals, for example, where some fixed costs (the machine costs or plant costs) cannot be changed in the short run. The other costs are called variable costs (e.g., labor costs).

The Firm Supply Curve under Perfect Competition

The cost curves we have reviewed can help to develop a theory of the supply curve for a firm, but to do so we must know something about the demand curve for the firm's product. In our earlier practice with supply and demand, the demand curve represented the market demand for the product. The demand for a single firm's product may be different. To gain an idea of what a typical firm's demand curve will look like, we must know what type of market structure we are talking about.

Several market structures provide insights to an idealized world or applicability to the real world. One defining principle that distinguishes the various market structures is the degree of control that individual firms have over the prices of their products. Two cases define the extreme forms of market structure: the competitive and the monopoly cases. We will look first at the competitive model, then discuss market structure more generally, and finally follow that discussion with the monopoly case.

The competitive model is rarely seen in the world in its idealized form. It requires several assumptions that ensure perfect competition. The assumptions are as follows:

1 A sufficient number of buyers and sellers of the good exist so that no single buyer or seller has any power over the price.
2 The good is homogeneous: that is, all producers produce the exact same good so that the market cannot be segmented on the basis of difference of goods.
3 Information is perfect. All buyers and sellers have information on all relevant variables such as prices and qualities.
4 No barriers to entry or exit are present. A producer starts producing, buying necessary machinery, patents, or anything else on terms that are equivalent to those already in the industry.

These assumptions ensure that a short-run market equilibrium can be represented by the price and the quantity at which demand and supply curves intersect. Figure 2.17 illustrates the model. Under the assumptions of competition, the demand curve facing the firm will be flat, as shown by the curve $D = MR = P$. To understand this point, consider a mental experiment. Suppose that the market for wheat was competitive and that it had determined, by the actions of market demand and supply, some equilibrium price for wheat, say \$3.50 per unit, as shown. Suppose now that a single firm chose to raise its price above \$3.50. Would anyone buy its product? They would not because they know (perfect information) that they can buy an identical product (homogeneous product) elsewhere for \$3.50. In theory, at even a slightly higher price, the quantity demanded would slide horizontally to zero. On the other hand, suppose that the farmer wished to double the output. Would the farmer have to lower the price in order to sell it all? This would not be necessary because the farmer's output is small relative to the whole market (numerous buyers and sellers); hence, the farmer could sell as much as he or she wanted at the going price of \$3.50.

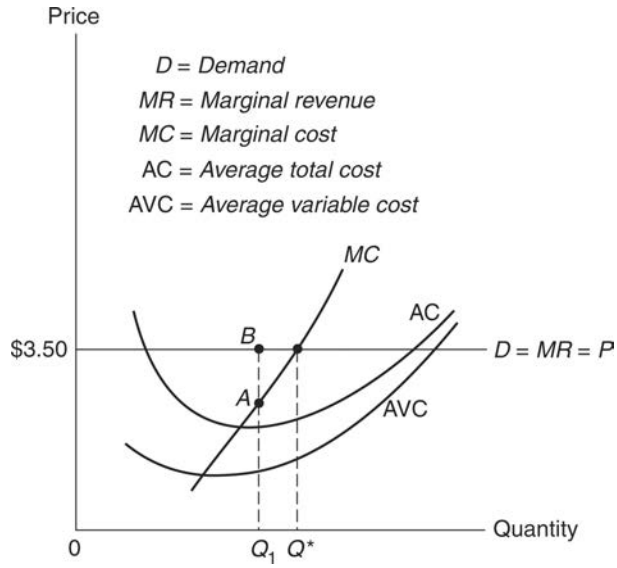


Figure 2.17 The Competitive Firm's Supply Curve

The demand curve for this firm is labeled $D = MR = P$ indicating that it represents demand, as well as the marginal revenue for the firm and the going market price. Note that the market price is identical to the firm's marginal revenue. Marginal revenue is defined as the extra revenue obtained by selling one more unit of product. Because this firm can sell all it wants at the going market price, it can sell the marginal unit at that price, as well. Thus, marginal revenue equals price in the competitive model. It is only in cases where a firm has some monopoly power that marginal revenue will differ from price.

The profit-maximizing output for this competitive firm can now be deduced. The firm will maximize its profits at that output where marginal revenue (the price) equals marginal cost; this occurs in the figure at output Q^*. This output is called the firm's profit-maximizing output. The common sense of this seemingly technical proposition can be understood by examining a "wrong" output level—one that is not profit maximizing—for example, Q_1. Suppose a firm that is currently producing Q_1 units were to produce one more unit. The cost of this one extra unit would be A, by definition the point on the marginal cost curve at $Q = Q_1$. The revenue from this one extra unit would be B (which equals the price), the point on the marginal revenue curve at $Q = Q_1$. The firm would increase its profits by producing that extra unit and would continue to increase its profits as long as the marginal revenue curve was above the marginal cost curve. Hence, maximum profits would occur only where $MR = MC$.

The supply curve for a firm shows the firm's profit-maximizing output at each possible price. The competitive firm is producing at the output where price equals marginal cost. If the market price were to rise in steps, the firm's adjustment steps would just be to follow the marginal cost curve on up.

The competitive firm's supply curve will be its marginal cost curve, as long as the price is sufficiently high to make it worthwhile to produce at all. Price must at least cover the firm's average variable cost (AVC).

The competitive market supply will be determined by the horizontal sum of the individual firm supply curves. This horizontal summing is done much in the manner in which we found market demand curves. The market supply curve in the competitive case, the sum of firm marginal cost curves, will also represent the industry marginal costs of production. In general, under competition, the supply curve is the industry marginal cost curve.

What then determines a good's market price? The answer is that combination of output and price at which market quantity demanded equals market quantity supplied, or the intersection of market demand and supply.

The assumption of free entry and exit, however, offers further insight into the workings of the competitive market. Suppose, for example, that the equilibrium price in the wheat market in the short run is high enough so that producers in the sector may earn attractive economic profits.[2] Any positive economic profit will be attractive to potential entrants. With perfect information and no barriers to entry, other suppliers will enter the market. This will increase market supply and drive down market prices. The entry process logically would continue in the long run until the prices have fallen enough to eliminate economic profits. In the long run, equilibrium profits will be zero, and price will be at the lowest point on each firm's long-run average cost curve.

However, if barriers to entry in the form of licensure or other restrictions exist, this adjustment process will be impeded. Sellers will be able to earn economic profits over long periods of time, perhaps indefinitely. Such a situation is not perfectly competitive even though the forces of demand and supply determine price at any moment. It is important to evaluate the extent to which all of the four conditions for competition stated at the beginning of this section are satisfied.

Monopoly and Other Market Structures

Unlike the perfect competitor, firms in other market structures have market power, which is the ability to affect market price. These market structures entail the pure monopoly (and the natural monopoly version), monopolistic competition, and the several forms of oligopoly. In nearly all instances, the market power concept determines the characteristic of choice of optimal price or quantity.

Consider the pure monopolist. A pure monopoly is an industry with a single seller who has no close substitutes. As such, the monopolist faces the whole market demand curve, which is usually downward sloping. Downward sloping demand and market power are synonymous concepts; this is because the monopolist, unlike the perfect competitor, will not lose all its customers when raising its price.

In health sectors, pharmaceutical firms that control patents for certain drugs may be pure monopolists. Individual physician practices are not pure monopolies, but because the numerous competitors of each are differentiated by reputation, patient loyalty, and patient/practice distance, each physician probably has some market power. Many economists treat physician markets as monopolistically competitive. Finally, when a town has only a few hospitals, each hospital also has some market power. There being few firms in that market, the hospitals would be considered an oligopoly. Equilibrium for the monopolist is illustrated in Figure 2.18. The demand curve facing the monopolistic firm is downward sloping because the monopolist faces the whole downward sloping market demand curve. With a downward sloping demand curve, the incremental or marginal revenue (MR) is less than the demand price. Why is this the case? Suppose the monopolist were selling Q_0 units at price P_0. Total revenue, TR_0, would be P_0Q_0. The monopolist would be selling to everyone who is willing to pay at least the price P_0. In order to sell one more unit of the good, the monopolist would have to induce more consumers to buy by lowering the price. It may be impossible to lower the price to extra consumers without also lowering to all previous consumers. In this case, because the monopolist must lower the price to everyone else, the marginal revenue will be

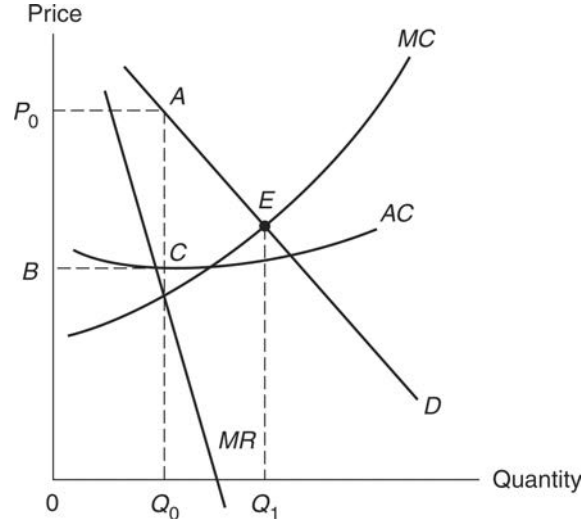


Figure 2.18 The Monopolist's Equilibrium

the price of the extra unit of the good sold minus the loss of revenue from everyone else who now pays less.

To maximize profit, the monopolist produces where $MC = MR$, at Q_0 in Figure 2.18. The corresponding price is P_0 and total profit is the rectangle P_0ACB. If barriers to entry are persistent, the economic profits can be maintained and even increased through advertising, promotion, new product development, or other means. The fact that the monopolist is earning excess profits suggests that the monopolist has reduced the amount produced from the competitive amount. The monopolist in the graph chooses point A on the demand curve. If the monopolist had acted like a competitor by setting a price equal to marginal cost (MC) it would have chosen quantity Q_1, point E, providing more output and charging a lower price. The induced scarcity caused by the monopolist necessarily raises the price to the consumer.

Economists seek to compare different economic situations or to examine the effects of different policies. In doing so, they often use the concept of allocative efficiency. One of the most widely used examples to illustrate the problem of inefficient resource allocation is found in the comparison of monopoly and competitive equilibria.

Consider once again the case of monopoly. Figure 2.19 shows the long-run marginal cost curve in an industry with constant marginal costs. With demand curve D, the competitive market equilibrium is at P_c and Q_c. Suppose, somehow, that the providers are able to form a monopoly. If so, it will be in their interest to raise prices by withholding services from the market. The resulting monopoly will produce at quantity Q_m, consistent with price P_m, where price is higher than marginal cost. The result reflects a loss to the consumer due to the monopoly. The total loss is indicated by the triangle ABC, called the welfare loss.

We can understand welfare loss better if we think in terms of marginals. Consider first that the demand curve measures the highest price that people are willing to pay for an extra unit of the good. The price they are willing to pay measures their marginal benefit. Now consider what the consumer and society as a whole have to give up when they face a

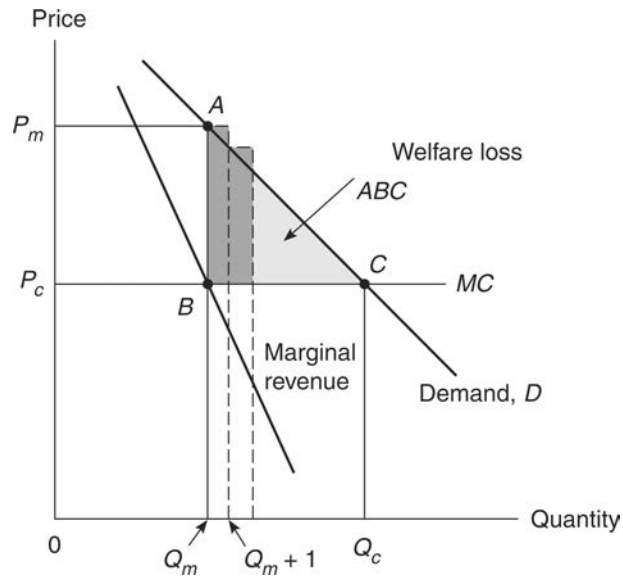


Figure 2.19 Welfare Loss of Monopoly

monopoly. The monopolist will choose output level Q_m. If we somehow could have produced one more unit of the good, the $(Q_m + 1)$th unit, we would have made a net gain for society. The benefit of that extra unit is $1 \times A$ in the graph because by marginal benefit (here equal to the height of the demand curve) we mean the benefit of the extra unit. Similarly, the cost to the monopolist, and thus to society as well, of the extra unit is $1 \times B$. Because the marginal benefit exceeds the marginal cost, the extra unit yields society a net gain of the rectangular shaded area.

BOX 2.2

Is Competition Better than Monopoly?

Virtually all economists greatly admire competition and competitive markets. But economists also understand that the benefits of competition are more likely to arise when competitive conditions are fully in place. The extent to which health care markets deviate from the perfectly competitive standard, and the effects of these deviations, are major themes in health economics (Gaynor and Town, 2012).

The 2010 passage of the Affordable Care Act (ACA) adds urgency to better understanding the role and effects of competition in health care delivery. Supporters of the ACA legislation often claim that it will promote competition among health insurers and make insurance more affordable. But will these predictions unfold? As health insurance exchanges were rolled out in the fall of 2013, the *New York Times* reported that 58 percent of the counties served by federal exchanges had only one or two insurers, i.e., these markets were highly concentrated. The *Times* article provides a vivid illustration for a 50-year-old who would pay $644 a month for coverage, before federal subsidies, in a rural Georgia county served by one insurer compared to $320 for comparable coverage in Atlanta, Georgia which was served by four insurance carriers.

Although rigorous scholarly studies of the ACA are not yet available, and there is some evidence of increasing competition in many markets since 2013, past research on the health insurance industry provides valuable clues. Daffny and colleagues examined employer-provided group plans over the period 1998–2006. These plans covered about 90 percent of the non-elderly with private insurance coverage. Most markets experienced an increase in concentration among insurers over the study period, with premiums rising about 7 percent in a typical market as a result of the increased concentration. The degree of monopoly is a recurring theme throughout the text in our discussions of efficiency and the role for regulation.

Sources: Reed Abelson, Katie Thomas, and Jo Craven McGinty, "Health Care Law Fails to Lower Prices for Rural Areas," *New York Times*, October 23, 2013 (nytimes.com/2013/10/24/business/health-law, accessed November 2016), Dafny, Duggan, and Ramanaraynan (2012), and Sheingold, Nguyen, and Chappel (2015).

Reasoning iteratively, another unit again yields another net gain to society, this time somewhat smaller than the first net gain. Net gains will continue to occur until we reach the output at which society's marginal benefit (demand) intersects its marginal costs, which occurs at *C*. The total net gain to society from increasing output up to the point at *C* equals the triangle labeled *ABC*.

Conclusions

The microeconomic tools developed in this chapter consist of the production possibility frontier, demand-and-supply analysis, utility and indifference curve analysis, production and cost curves of a typical firm, firm behavior under competition and monopoly, and the measure of welfare loss. The economic tools used later in the text apply or extend the tools developed here. By learning these ideas, you will gain an understanding of the terminology used in health economics, as well as an understanding of the type of reasoning used.

Summary

1 The concept of scarcity underlies much economic thinking. Scarcity necessitates that decision makers make trade-off decisions at the margin. The production possibilities frontier represents these trade-offs, and its slope represents the opportunity cost of one good in terms of the other.
2 Supply-and-demand analysis of competitive markets is a basic tool of economics and provides insights that extend beyond the theoretical, perfectly competitive markets. Supply reflects sellers' offers as a function of price, and demand reflects buyers' offers as a function of price. The intersection of demand and supply describes the market equilibrium.
3 Comparative static analysis of demand and supply finds the new equilibrium after economic events shift either curve. Demand-increasing (-decreasing) events tend to raise (lower) equilibrium price, while supply-increasing (-decreasing) events tend to lower (raise) equilibrium price.
4 A function describes a relationship between one or more independent variables yielding a unique value for the dependent variable. The linear demand function, showing demand as a straight line, is only one special case of the many possibilities.
5 The utility function summarizes a consumer's preferences. Higher utility numbers are assigned to consumer bundles that provide higher levels of satisfaction, meaning that the consumer prefers these bundles.
6 Indifference curves describe bundles of goods that yield the same utility and hence the same level of satisfaction. Well-behaved indifference curves are downward sloping, continuous, and convex to the origin.
7 The budget constraint represents the combinations of goods that the consumer can afford given his or her budget. The budget constraint is downward sloping, and its slope is the negative of the ratio of prices.
8 In consumer theory, the consumer maximizes utility subject to a budget constraint. This means that the consumer picks the most preferred consumer bundle from among those he or she can afford. The equilibrium occurs at the tangency between the budget constraint and the highest attainable indifference curve.
9 Price elasticity depicts the responsiveness of demand to changes in price. It is defined as the ratio of the percentage change in quantity demanded to the percentage change in price. Each other elasticity also represents the ratio of a percentage change in a dependent variable to a percentage change in a given independent variable.
10 The production function describes the relationship of inputs to output. The marginal product of an input is the increase in output due to a one-unit increase in the input holding all others constant. That marginal product tends to decline as more input is added describes the law of diminishing marginal returns.

11 The average total cost curve of a firm shows the total cost per unit of output. The marginal cost curve shows the extra cost required to produce an additional unit of output.
12 The competitive firm in the short run produces that output where price equals marginal cost. The marginal cost curve is therefore the supply curve of the competitive firm.
13 In long-run equilibrium, entry by competing firms forces the typical competitive firm to produce an output level such that its price equals its minimum average cost. At this output, the competitive firm is producing the economically efficient output, and it is earning zero economic profits.
14 The pure monopolist faces the entire downward sloping market demand curve, and this implies that its marginal revenue lies below the demand curve. The monopolist restricts output, by comparison to the competitive case, and it charges a higher market price.
15 The pure monopoly case is one instance of a market in which a welfare loss occurs. A welfare loss, represented by an area under the demand curve and above the marginal cost curve, is an opportunity for mutual gains that is being foregone by the market.

Discussion Questions

1 Explain the difference between cardinal and ordinal utility. Do you think that it is possible for researchers to find out which type of utility people actually have?
2 If a consumer always prefers more to less, can the indifference curves between the two goods be upward sloping? Explain. What if one of the "goods" is actually something unpleasant, like broccoli to a three-year-old?
3 The law of diminishing marginal returns states that eventually the marginal product of an input will tend to fall as more input is added. Describe real-life scenarios, explaining why this is likely to happen. For example, imagine a backyard garden of fixed size and all other inputs except labor also fixed; will adding a worker increase your output? Will adding another increase output by as much? Another?
4 If it makes sense that one type of labor can substitute for another in production, how can capital, a physical object, substitute for labor, a human being?
5 Describe the long-run equilibrium of the competitive firm. Conceptually remove a single assumption of perfect competition and analyze whether and how the process of long-run equilibrium would change. For example, if information were very imperfect, would the long-run equilibrium be achieved? If the firms' products were not exactly alike? If there were barriers to the entry of new competitors?
6 When a welfare loss occurs because of monopoly, what exactly is lost? Who loses it?
7 Resolve the following: "The price of ice cream increased in the summer, yet quantity also increased. Therefore, the law of demand does not apply to ice cream."
8 How is a production function affected by the invention of a new process related to it? Can this change result in lower prices to the consumer? What do you think? Do improvements in technological knowledge in the production of consumer goods necessarily reduce average family expenditures?

Exercises

1 Draw a production possibilities curve for an economy that produces two goods, health and entertainment. Show how this PPF would change if the technology for improving

peoples' health was to improve. Show the change if there were an increase in the underlying available quantities of capital and labor.

2 Draw a production possibilities curve between health and all other goods. Insert a point in the drawing that illustrates an economy with an inefficient health system. Insert two additional points that illustrate two efficient economies but two that contrast in their relative emphasis on health care versus all other goods. Is there a cost to society of policies that lead to increases in health care? Explain.

3 Using a supply-and-demand graph and assuming competitive markets, show and explain the effect on equilibrium price and quantity of the following:
   (a) A technological change that reduces the cost of producing X-rays on the market for physician clinic services.
   (b) Increased graduations of new doctors on the market for physician services.
   (c) The virtual elimination of smoking in the population on the market for hospital services.
   (d) A price ceiling placed on physician fees in the market for physician services.

4 Graph the following demand functions:
   (a) $Q_d = 110 - 3.3P$
   (b) $Q_d = 100P^{-1.3}$

5 In a graph with OG on the vertical axis and Food on the horizontal, what is the Food-axis intercept? How does the budget constraint shift if the consumer's income level and the two prices all double?

6 Calculate the price elasticity for a $1 change in price at initial price level $300 for the demand function $Qd = 1{,}500 - 1.5P$.

7 What is the slope of the isoquant described by the data in Table 2.5 when evaluated from a labor input of 7 to 8?

8 Assume that a monopoly firm has a linear demand curve and a constant marginal cost curve. Graph this firm's optimal output choice before and after a per-unit excise tax is placed on the output. Does the equilibrium price rise by as much as the tax?

9 Using equation (2.4), what is the demand equation as a function of P_s if the price of other pastas (P_o) is $3, the individual's income (Y) in thousands is $25, and tastes ($Z$) are represented by 20? What happens if the individual's income increases to $40?

10 Colorado legalized the recreational use of marijuana in 2014 to raise revenues through sales and excise taxes. In Denver, 1/8 ounce of marijuana costing $30 had about $8 in various taxes added on. Assuming a price elasticity of demand for marijuana of –0.7 (see Table 2.3), determine the percentage change in quantity that would result from a price increase of $8. Under what circumstances would prices rise by less than the full amount of a tax? How does the price elasticity of demand affect tax revenues?

Notes

1 For example, it becomes possible to compare the price elasticity of demand for beef with that for automobiles, even though the price levels and quantities are different.

2 Economic profits represent profits after considering all costs including opportunity costs. A "normal" level of profits is necessary to keep firms in the market, and is considered a factor payment to the entrepreneur, just like wage and salary payments to workers or rents paid to owners of buildings and machinery.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 3

Statistical Tools for Health Economics

In this chapter

- Hypothesis Testing
- Difference of Means
- Regression Analysis
- Multiple Regression Analysis
- Statistical Inference in the Sciences and Social Sciences
- Conclusions

The ideas from economic theory must be tested and measured according to the standards of real-world data. Statistical techniques applied to economics are collectively called econometrics. In Chapter 2, we discussed supply and demand, as well as the importance of price and income elasticities. Economic theory predicts that demand curves will slope downward, but it does not predict the degree of responsiveness of demand to price and other variables; it is the task of statistical analysis to estimate these magnitudes. When close substitutes are available for a good, theory predicts that demand will be more sensitive to price than if no close substitutes are available. Yet it is hard to know whether a 1 percent increase in price will decrease the quantity demanded by 10 percent, 1 percent, or 1/10 of a percent, yielding elasticities of –10.0, –1.0, or –0.1, respectively. Measurements of the economic parameters may prove crucial in analyzing whether drug companies raise drug prices, whether higher insurance copayments will lead people to use less treatment, or whether mandated levels of health care are economically efficient.

This chapter considers statistical methods that econometricians use to draw inferences from data that are collected. Many students with natural science backgrounds are familiar with laboratory experiments which seek to hold the environment constant and administer treatments to experimental groups. Researchers then compare the results to those of untreated control groups. One form of this design is the *dose-response* model in which researchers relate the results or responses to the experimental treatment, or the dose. If statisticians determine that the resulting differences are *significant*, a term we will discuss in this chapter, they consider the dose to be effective.

Social science analysts of human behavior rarely find an experimental group to match with a convenient control group. Instead, they must usually collect information from people doing day-to-day activities. Using statistical methods, they try to control for the confounding differences among the people that they are analyzing. The more successful they are in controlling for such differences, the more reliable the analysis will be.

This chapter begins with discussions on how we form hypotheses. It then considers difference of means analysis as a way of introducing statistical inference. Most of the rest of the chapter concentrates on the simple and multiple regression analyses that economists most often use.

Hypothesis Testing

Economists studying health care often face statements that, while plausible, demand some validation:

"Men and women don't smoke the same numbers of cigarettes."
"Rich people spend more on health care than do poor people."
"The United States spends more on health care than does the United Kingdom."

Either logic or casual observation would suggest that these statements are true. It would be useful, however, to have a rigorous method of determining whether the assertions are correct. Statistical methods suggest formulating these statements as hypotheses and collecting data to determine whether they are correct.

Take, for example, the first assertion about smoking levels. We state clearly both the hypothesis we wish to disprove (called the null hypothesis), as well as the hypothesis the theory suggests to be the case (the alternative hypothesis). The null hypothesis here, H_0, is that men's levels (c_m) equal women's levels (c_w), or

$$H_0 : c_m = c_w \tag{3.1}$$

The alternative hypothesis H_1, is that c_m does not equal c_w:

$$H_1 : c_m \neq c_w \tag{3.2}$$

We seek convincing evidence that c_m differs from c_w. Hypotheses that are designed to test for equality among two or more items are sometimes called *simple* hypotheses.

Consider the second hypothesis, which asserts that rich people spend more on health care than do poor people. If we define health care expenditures of the rich as E_r and the poor as E_p, then the null hypothesis is:

$$H_0 : E_r = E_p \tag{3.3}$$

The alternative is:

$$H_1 : E_r > E_p \tag{3.4}$$

In this analysis, it may not be enough just to show that E_r differs from E_p. Even convincing evidence that the poor spend more, or $E_p > E_r$, would not validate the hypothesis. Hypotheses that test whether two or more items are greater (or less) than each other are *composite* hypotheses. Having seen how one might construct the hypotheses in question, we now discuss how to test them.

Difference of Means

In the past decade, health professionals have turned their attention, often with some alarm, to the increased level of population obesity. According to the Centers for Disease Control, obesity is "common, serious and costly" with:

- More than one-third of U.S. adults (35.7%) obese.
- Obesity-related conditions including heart disease, strokes, type 2 diabetes and certain types of cancer, some of the leading causes of preventable death.
- An estimated annual medical cost of obesity in the U.S. of $147 billion in 2008 U.S. dollars; the medical costs for people who are obese were $1,429 higher than those of normal weight.

Source: *Adult Obesity Facts*, www.cdc.gov/obesity/data/adult.html, accessed February 14, 2014.

Although we will look more explicitly at obesity in Chapter 7, measuring its incidence is an important and useful topic in the use of quantitative methods.

The body-mass index or BMI is probably the most common (although far from the only) measure of obesity. Researchers calculate the BMI by dividing the subject's weight in kilograms by height in meters squared. The formula then is:

$$BMI = \frac{mass}{height^2}$$

So Pat, a research subject with a weight of 170 pounds (77.1 kilograms) and a height of 66 inches (1.67 meters) will have a BMI of 27.6. Medical and public health experts consider a BMI over 25 as "overweight," and one over 30 as "obese." According to the BMI, Pat is overweight.

Suppose we wish to compare the BMI of men with that of women. We can weigh and measure one man and one woman, and find the man's BMI of (B_M) 27 and the woman's BMI (B_W) of 25. This provides evidence that men have higher BMIs than women, because $B_M > B_W$, or 27 > 25. It is not very convincing evidence, however. The man or woman, or either, may not be typical of the entire group. What if a different man and/or woman had been selected? Would the answer have been different?

It seems logical to test several men and to compute the mean or average level by summing the measurements and dividing them by the total number of men tested. The National Institutes of Health (in 2001 and 2002) collected a database of over 43,000 people ages 18 and over, called the National Epidemiologic Survey on Alcohol and Related Conditions, or NESARC. Although this sample focused on potentially substance abusive activities including smoking, drinking of alcoholic beverages, and the taking of recreational (and harder) drugs, it also measured height and weight. The textbook authors found that

For 18,159 men, the mean, or average level, $\overline{B}_M$ was 27.19
For 23,495 women, the average level, $\overline{B}_W$, was 26.33.

The difference, $d = \overline{B}_M - \overline{B}_W$ then, is 0.85, evidence that men have higher BMIs which indicate the problem of obesity.

The Variance of a Distribution

Although a difference of the two means is improved evidence, the econometrician desires a more rigorous criterion. Suppose the true BMI for both men and women is 27.00 but our sample randomly drew a higher average level for men (27.19) than for women (26.33). Figure 3.1 plots the distributions in percentage terms. The women have higher percentages in the lower categories and the higher categories. For example, 39.45 percent of the women have BMI between 20 and 25, in contrast to only 31.07 percent of the men. However, men are far

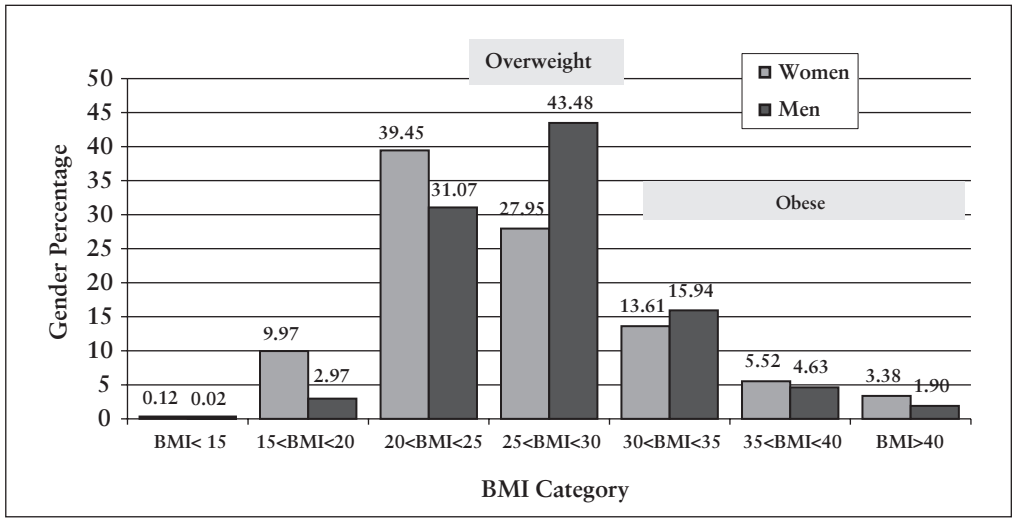


Figure 3.1 BMI by Gender, 2001–2002

more likely to be overweight (BMI between 25 and 30); 43.48 percent of the men have BMI between 25 and 30, compared with 27.95 percent of the women.

Statisticians have found the variance of a distribution to be a useful way to summarize its dispersion. To calculate the variance of women's levels, we subtract each observation from the mean (26.33), square that term, sum the total, and divide that total by the number of observations, *N*. Hence, for the 23,495 women in the sample, variance, V_w, equals:

$$V_w = \frac{N_{\min} \times (10.98 - 26.33)^2 + N_2 \times (BMI_2 - 26.33)^2 \ldots + N_{\max} \times (78.12 - 26.33)^2}{23,495} \quad (3.5)$$

Here $N_{\min}$ is the number of women with the minimum calculated BMI, N_2 is the number with the next level of BMI, and so on (and yes, there was a woman with a BMI of 10.98 and another with a BMI of 78.12!).

V_w reflects the variance of any individual term in the distribution. If *V* is large, then the dispersion around the mean is wide and another woman tested might be far from our mean. If *V* is small, then the dispersion around the mean is narrow and another observation might be close to the mean.

Standard Error of the Mean

The variance is often deflated by taking the square root to get the standard deviation, *s*, yielding:

$$s_w = \sqrt{\frac{N_{\min} \times (10.98 - 26.33)^2 + N_2 \times (BMI_2 - 26.33)^2 \ldots + N_{\max} \times (78.12 - 26.33)^2}{103,444,772}} \quad (3.6)$$

As with *V*, a large (small) value of *s* indicates a large (small) dispersion around the mean. Statisticians have shown that we can calculate the standard error of the mean itself by dividing *s* by the square root of the number of observations, and because the data came from a population weighted sample (the 23,495 women were statistically chosen to represent all women over the age of 18), the denominator is over 100 million (women). In this sample, the standard deviation of the distribution for women equals 403.07—the average woman may vary greatly from the mean. The standard error of the mean of the women's distribution would then equal s_w divided by the square root of 103,444,772, or (403.07 ÷ 10,171), which equals 0.04.[1]

A powerful theorem in statistics, the Central Limit Theorem, states that no matter what the underlying distribution, the *means* of that distribution are distributed like a normal, or bell-shaped, curve. Hence, we can plot the normal distribution of means of women's levels with a mean of 26.33 and a standard error of 0.04.

Statisticians have also shown that a little more than 68 percent of the area under the curve would be within one standard error, or between levels of 26.29 (that is, 13.47–0.04) and 26.37 (i.e., 13.47 + 0.04). About 95.4 percent would be within two standard errors. This means that we could be about 95 percent sure that the true BMI for the women over the age of 18 is between 26.25 and 26.41. A similar calculation can be done for men, yielding a similar measurement. Intuitively, the further apart the means and the smaller the dispersions (standard errors), the more likely we are to determine that the average level for men is greater than that for women. To test the hypothesis formally, we then construct a "difference of means" test. We wish to compare measurement $d = \overline{BMI}_m - \overline{BMI}_w$, to zero, which was the original hypothesis.

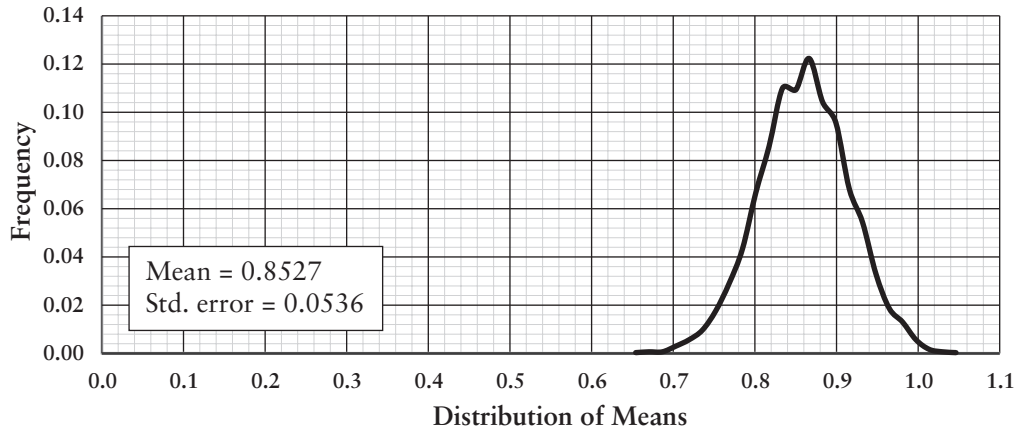


Figure 3.2 Difference of Means Test for BMI

Here $d = 0.8527$. The variance of the difference is defined as the sum of the variances of the standard errors. The standard error for women was 0.0396 as we calculated it, and the standard error for men was 0.0362, so the standard error of the difference would be:

$$s_d = \sqrt{0.0396^2 + 0.0362^2} = 0.0536 \tag{3.7}$$

We plot the difference and its distribution in Figure 3.2.

The most probable value of the difference, as noted in Figure 3.2, is 0.8527. About 68 percent of the distribution lies between 0.7991 (i.e., 0.8527 – 1 × 0.0536) and 0.9064 (similarly, 0.8527 + 1 × 0.0536). About 95.4 percent of the distribution lies within two standard deviations, between 0.7454 and 0.9600. This experiment would find very good evidence that women have a lower BMI than do men. The probability is well over 99 percent that this difference is statistically significant, that is not equal to 0.

Alternatively, the *t*-statistic, comparing the numbers 0.8527 and 0.0, equals 0.8527 ÷ 0.0536, or approximately 15.90. Statisticians calculate tables of *t*-statistics, whose critical values are related to the size of the sample. With an effective sample over 200,000,000, a *t*-statistic of nearly 16 is statistically significant at well over the 99 percent level. In other words, we can be 99+ percent certain that men have a higher BMI than do women.

Hypotheses and Inferences

This process illustrates the steps that are necessary to test hypotheses appropriately. The econometrician must:

1 State the hypotheses clearly

$H_0 : BMI_m = BMI_w$, against
$H_1 : BMI_m \neq BMI_w$.

2 Choose a sample that is suitable to the task of testing.

3 Calculate the appropriate measures of central tendency and dispersion: the mean and the standard error of the mean for both men and women, leading to the difference of the two means.
4 Draw the appropriate inferences: men smoke more than women.

No matter how sophisticated the method used, good statistical analysis depends on the ability to address these four criteria and stands (or falls) on the success in fulfilling them. Box 3.1 provides an interesting example of how analysts have examined the potentially cancer-causing properties of cellular phones.

There are, of course, measures of central tendency other than the mean (or average). Someone with a BMI of 50 may unduly influence the mean. A different measure, the median, calculates a statistic such that half of the observations are greater than the median and half are less. Thus, a median BMI of 26 would imply that half of the people had a BMI of less than 26, and half more. The median is less sensitive to extreme values in the data (e.g., someone with a BMI of 50 has no more impact on the median than someone with a BMI of 30). However, the median can present mathematical problems in hypothesis testing. Simple formulas for standard errors of medians have not been available, although popular numerical "bootstrapping" methods now provide intuitive and accurate standard errors. For a good discussion of bootstrapping, see Efron and Tibshirani (1993).

BOX 3.1

Do Cell Phones Cause Cancer?—Positive Reports but Inconsistent Data

In 2011 the International Agency for Research on Cancer (IARC) publicized its *Monograph on the Evaluation of Carcinogenic Risks to Humans*. The monograph classified exposure to mobile phones as category 2B, "possibly carcinogenic to humans." In a *National Cancer Conversation Series*, Dr. Martha Linet, a long-time researcher at the National Cancer Institute, spoke about this classification.

Is there evidence of an increased risk of cancer from mobile phone use?

> [Dr. Linet responded] that while most studies to date have not found an association between cell phone use overall and the development of tumors, "a handful of studies" have shown an association with increased risk for glioma (a type of brain cancer) among the small number of cell phone users who reported the highest level of call time. Among the positive studies there were conflicting results, and they did not show increasing incidence as the exposure (to cell phones) increased. She also said there was "no biologically plausible mechanism or animal evidence for how cell phones might cause cancer."

Why were there inconsistencies among the studies?

> Most of the studies were interview-based from brain tumor patients and control subjects. Cell phone technology has changed over the years, as has the way people

use cell phones. Both of these could make accurate recall difficult. Further, none of the epidemiologic studies measured actual radiofrequency exposure to the brain.

Readers should note that other substances having classification 2B include lead, coffee, and pickled vegetables (a traditional Asian dish). The "takeaway" said Dr. Linet, is that the term *possible carcinogen* (emphasis added) is appropriate if "one keeps in mind that possible means 'maybe,'" relating to "positive reports but overall inconsistent data."

Sources: IARC (2013). Dr. Martha Linet on Cell Phone Use and Cancer Risk—NCI Cancer Bulletin for June 28, 2011—National Cancer Institute, http://elbiruniblogspotcom.blogspot.com/2011/06/dr-martha-linet-on-cell-phone-use-and.html, accessed July 1, 2016.

Regression Analysis

The difference of means analysis is extremely useful in treating continuous data that can be broken up by categories, such as gender, race, or location.[2] Yet many interesting economic variables occur naturally as continuous variables. Health care expenses, physician visits, firm profits, as well as prices and/or incomes could take large numbers of values naturally, and if we group them into categories we lose considerable information. If we have information on income, in terms of dollars per year, we can distinguish among households with incomes of $10,000, $20,000, $30,000, and so on. If we defined high income as greater than $100,000, for example, separating all of the different incomes into two categories, we would then have no way of distinguishing between households with incomes of $10,000 and $20,000 (or, for that matter, between households with incomes of $125,000 and $250,000).

Regression analysis allows the econometrician to fit a straight line through a set of data points. How might we use this for analyses? We have seen that men seem to have higher BMIs than women. Older people may have different BMIs than do younger people. African-Americans may differ from whites or Hispanics either due to their incomes or their tastes for different kinds of foods. Smokers or drinkers may have higher (or lower) BMIs. High income people may eat more (raising the BMI), but they may eat healthier foods (lowering the BMI).

Ordinary Least Squares (OLS) Regressions

Econometricians use two rules to determine this line. The first rule is that the deviations (unless the line fits perfectly) from the line must sum to zero. Positive deviations must be offset by negative deviations. Many lines, however (see Figure 3.3), have this characteristic (for example, dashed lines R_1 and R_2). It is necessary to have a second criterion by which to distinguish among the large number of lines where the sum of the deviations equals zero.

The second criterion minimizes the sum of the squared deviations of the actual data points from the fitted line. Even though the sum of the deviations equals zero, the sum of the squared deviations must be positive (any number multiplied by itself is either zero or positive). Hence, one can choose among the many lines with sums of zero deviations by picking the single line with the minimum or least sum of the squared deviations. Such analyses are *ordinary least squares* (OLS) analyses.

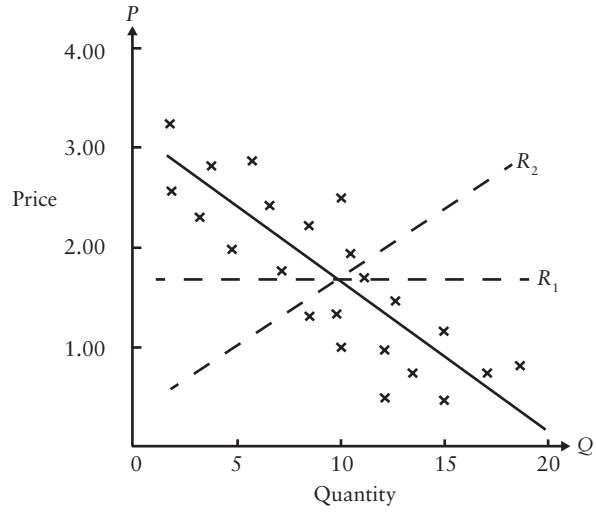


Figure 3.3 Plot of Quantity Purchased against Price

Suppose, for example, we wish to relate people's BMI, for example, to their incomes. The resulting equation would have the following form:

$$BMI = a + b\ Income + \varepsilon \tag{3.8}$$

where a and b are the parameters to be estimated. Parameter a is sometimes referred to as the constant, or the intercept.

Parameter b refers to the slope of the line and shows the direction and magnitude of the impact of a change in income, for example, on the BMI. If a higher income level has a positive value of b, it implies that more affluent people are eating more, and getting fatter. If b is negative, it implies that affluent people are eating better food that lowers their BMI.

The last parameter is the error term ε. No regression analysis will fit the data exactly. Errors are likely and may reflect several causes. We may have omitted a variable, such as education (more educated people may be more aware of the dangers of obesity and eat less). We may have measured one or more of the explanatory variables, or the dependent variable (the BMI itself—this could happen if the data are self-reported, and people overestimate their heights and underestimate their weights), inaccurately. All of these may stand in the way of our predicting the amount demanded exactly. In advanced econometric work, understanding ε is crucial for ensuring that the estimated parameters are accurate. Our exposition here will assume that ε obeys the rules to allow us to make appropriate inferences with OLS analyses. We will examine some exceptions later in the chapter.

A Simple Regression

Here is a simple regression of BMI on income in thousands of dollars.

$$BMI = 26.709 + 0.00129 \times \text{income (in \$1,000)}, R^2 = 0.0001$$
$$(0.00060) \tag{3.9}$$

This equation indicates that a $1,000 increase in income is correlated with a very small (0.00129) increase in BMI. The *standard error of estimate* for the coefficient of income is

0.00060. This term is similar to the standard error of the estimated mean in the example of BMI for men and women earlier in the chapter. As before, the smaller the standard error, relative to the estimated value of b (in this case, 0.00129), the better the estimate. In this regression, the standard error of 0.00060 is relatively small compared to the coefficient, 0.00129; hence, the coefficient is significantly different from zero. The expression R^2 measures the fraction of the variation of the BMI explained by income alone. An R^2 of 0.0001 implies that variation in income explained one-hundredth of 1 percent of the variation in BMI. In short, income alone explains some, but not very much of the variation in obesity, and its incremental impact is tiny.

It is useful to examine this simple regression in detail because it has many features that occur in more complex analyses. Consider the following hypothesis:

H_0: Income doesn't matter; that is, $b = 0$ against the alternative hypothesis, which is:

H_1: Income is positively related to BMI; that is, $b > 0$. The test of the hypothesis is similar to a difference of means test. In particular, we are testing the difference between 0.00129 (estimated with standard error 0.00060) and 0.

The *t*-statistic here is 2.15; that is, the value of the coefficient, 0.00129, divided by the standard error of 0.00060. The value of 2.15 suggests that we can be more than 95 percent sure that the income has an increasing impact on BMI. This term is statistically different from zero.

If we have explained less than 1 percent of the variation of the BMI, then well over 99 percent is unexplained. In part, this occurs because the regression does not include some variables that are likely to be important. We noted earlier that several other variables might help explain more of the variation in BMI. The inclusion of more variables in a multiple regression is explained later.

This example also illustrates cross-sectional analysis, which provides snapshots of a slice of the population at one period in time. Because 2001–2002 was the first time that the NESARC was collected, it could not yet follow the people in the sample over time, and researchers could not be aware of continuing health problems, changes in wealth or income, or *systematic* differences in health or ability (that cannot be measured), that cross-sectional models treat as "random noise." As a result, cross-sectional regressions often explain less variance than panel data, which follow units of observation (individuals or households) over time, or time-series data, which calculate aggregates over time.

Estimating Elasticities

Analysts also use regressions to estimate elasticities. The definition of the income elasticity of demand (E_y) is the percentage change in quantity demanded, elicited by a 1 percent change in income and written as:

$$E_y = \frac{\%\text{ change in BMI}}{\%\text{ change in income}} = \frac{\left(\frac{\Delta BMI}{BMI}\right)}{\left(\frac{\Delta Y}{Y}\right)}, \text{ or } E_y = \left(\frac{\Delta BMI}{\Delta Y}\right)\left(\frac{Y}{BMI}\right)$$

ΔY stands for a one-unit (here \$1,000) change in the income, while ΔBMI represents the resulting change in the BMI. It follows that $\Delta BMI/BMI$ is a measure of the percentage change in quantity, whereas $\Delta Y/Y$ is a measure of the percentage change in income. In rearranging terms at the right, the term $\Delta BMI/\Delta Y$ represents the ratio of changes. With the linear function here, this is 0.00129.

In calculating an elasticity from coefficients estimated in a regression, we could find a different elasticity at any given income. Therefore, we want reference values for *BMI* and *Y*, and the mean (or average) values are often used. In our sample, the mean BMI is 26.75, and the mean income (in dollars × 1,000) is 29.86.

Hence, calculated at the mean,

$$E_y = 0.00129 \times (29.86 \div 26.75), \text{ or } 0.00144 \tag{3.10}$$

This says that a 10 percent increase in the income relates to (and possibly causes) a 0.14 percent increase in BMI. Does this make sense? Certainly at lower income levels (close to starvation, perhaps), increased income allows people to buy food to live on and fight malnutrition. However, at higher income levels, it is possible that the relationship could change.

Multiple Regression Analysis

Real-world relationships are seldom two-dimensional, as useful as this situation would be in drawing graphs. BMI may differ by gender, race, and ethnicity, and it may also relate to age and lifestyle variables. Some variables which are crucially important to determining BMI, such as basal metabolic rate (the rate at which the body metabolizes nutrition), may not be available outside of a clinical setting. Indeed, the omission of important variables may lead to particular behavior in the error term ε.

If we could graph each relationship, assuming that nothing else was changing, then simple regression would work fine. Fortunately, however, the mathematics necessary to estimate the appropriate relationship can easily accommodate more than two dimensions. We write the following multiple regression:

$$BMI = a + bY + cG + dA + eE + fO + \varepsilon \tag{3.11}$$

where:

Y = income
G = gender (male or female)
A = age
E = ethnicity
O = all other variables, including health and lifestyle variables

Although the example presented in Table 3.1 will now summarize numerous dimensions rather than original two dimensions, we use exactly the same least-squares criteria as before. The interpretation of the coefficients is similar to before but we can do it with more confidence. With the simple regression, relating *BMI* only to *Y*, the econometrician would not know whether gender, or age, or lifestyle, was varying as well. Including them in this regression allows us to "hold constant" these other variables and reduce the error. As a result, effects can now be calculated under the condition that "all else is equal." The R^2 measure of variation explained earlier also is available here.

R^2 will always rise with more variables (adding additional variables cannot explain *less* of the variation!). Several methods can be used to interpret R^2, and some statisticians wish to maximize R^2: that is, to explain as much variation as possible.[3] While this may be desirable, most econometricians are at least as interested in the values of the parameters estimated.

Interpreting Regression Coefficients

Table 3.1a shows the result of a multivariate regression of BMI against a number of factors that might affect it.

$$BMI = 17.51 + 0.90 * male + 1.61 * AfAm + 0.29 * Age - 0.0029 * Age^2 + 0.94 * Health + 0.15 * Spouse + 0.59 * Hispanic - 0.00055 * Income; R^2 = 0.0864 \quad (3.12)$$

This equation verifies the earlier finding from the simple comparisons, that men are more likely to be obese, but it also indicates that African-Americans, people with spouses, and Hispanics are all likely to have higher BMIs. Health is an index (where 1 is good, and 5 is bad), so bad health is correlated with a higher BMI. The effect of age is non-linear. Evaluating the coefficients, people tend to have increasing BMIs up to age 50, but they fall off slightly after that. Income, which showed a slight positive impact in the simple regression, now shows a slight negative, but insignificant, impact here. All in all, the regression explains about 8.6 percent of the BMI.

Table 3.1 Economic Determinants of BMI

Determinants of BMI	*(a)*			*(b)**		
	Coeff	*Std.Error*	*t-test*	*Coeff*	*Std.Error*	*t-test*
Intercept	17.50638	0.18327	95.52	18.02081	0.22234	81.05
(0 if female; 1 if male)	0.89674	0.05311	16.88	1.27763	0.05721	22.33
(1 if African-Am; 0 otherwise)	1.61420	0.08288	19.48	2.51273	0.11202	22.43
Age	0.29164	0.00811	35.96	0.28948	0.00817	35.42
Age squared	–0.00290	8.02E-05	–36.20	–0.00295	8.09E-05	–36.44
Health status (1 = excellent – 5 = poor)	0.93704	0.02464	38.03	0.94584	0.02513	37.64
(1 if spouse; 0 otherwise)	0.14897	0.05566	2.68	0.08652	0.05559	1.56
(1 if Hispanic; 0 otherwise)	0.59101	0.08263	7.15	0.60297	0.08515	7.08
(1 if African-Am+male; 0 otherwise)				–2.18621	0.16274	–13.43
Income x 1,000	–0.000552	0.00061	–0.91	–0.000557	0.00061	–0.92
1 if lifetime alcohol abstainer				–0.61792	0.08068	–7.66
1 if monthly drinker				–0.59014	0.05964	–9.89
1 if binge drinker				–0.01183	0.01516	–0.78

Table 3.1 *continued*

Determinants of BMI	*(a)*			*(b)**		
	Coeff	*Std.Error*	*t-test*	*Coeff*	*Std.Error*	*t-test*
1 if uses addictive drugs				–0.25827	0.11421	–2.26
1 if previously used addictive drugs				–0.01750	0.07381	–0.24
1 if "ever smoked"				0.19791	0.07077	2.80
1 if "currently smokes"				–0.88209	0.07724	–11.42
R^2	0.0864			0.0986		

Note: * Regression contains geographic regions, which are statistically significant.
Source: Computations from National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) by A. Goodman, 2014.

Regression (b) adds some interesting covariates relating to "lifestyle" behaviors. It turns out that both lifetime alcohol abstainers, as well as those who drink monthly, have BMI values about 0.6 less than others. Current users of addictive drugs have BMIs that are about –0.25 points lower than others. Current smokers have BMIs about –0.88 points less (smoking apparently does keep the weight down), while ex-smokers have BMI values about 0.2 points higher. Income, which had seemed to be slightly positive in its impact in earlier regressions, has a slightly negative and statistically insignificant impact, in the full regression.

Most often, econometricians are interested in whether coefficients are positive or negative and whether they differ significantly from zero. In a now famous example, Box 3.2 presents the surprising results of a detailed multivariate analysis looking at the impacts of hormone replacement therapy on postmenopausal women.

BOX 3.2

Hormone Replacement Therapy—Rigorous Statistics Reveal Surprising Results

As of July 2002, approximately 38 percent of postmenopausal women in the United States used hormone replacement therapy (HRT). While the U.S. Food and Drug Administration-approved indications for HRT included relief of menopausal symptoms (hot flashes, night sweats, and vaginal dryness) and prevention of osteoporosis, long-term use had been common to prevent a range of chronic conditions, especially heart disease. Advertisements by drug companies urged women to take HRT so they would stay "forever feminine."

Many scientists had expressed concern that studies finding benefits of HRT were based on nonrandom samples of women who sought out the hormone therapy. Users of HRT were better educated and healthier than postmenopausal women who did not take HRT. Thus, some researchers felt that "selection bias" could account for the effectiveness of HRT because those women for whom it was not effective and those who found the side effects bothersome or harmful, as well as less educated and sicker women, were not included in the studies.

Between 1993 and 1998, a randomized clinical trial called the Women's Health Initiative (WHI) studied 16,608 postmenopausal women aged 50 to 79. The study randomly assigned roughly half of the participants to the experimental group receiving HRT, a daily tablet containing conjugated equine estrogen and medroxy progesterone acetate (progestin). It assigned the other half to the control group receiving a placebo (an inert pill with no medical properties). Study participants were contacted by telephone six weeks after randomization to assess symptoms and reinforce adherence. Follow-up for clinical events occurred every six months with annual in-clinic visits required.

Researchers measured a multitude of health outcomes related to cardiovascular disease, stroke, cancer, fractures, and death. The complex statistical analysis compared health outcomes for the experimental group who took the estrogen/progestin tablet to the control group who took the placebo. Formal monitoring began in the fall of 1997, with the expectation of final analysis in 2005 after an average of approximately 8.5 years of follow-up. An independent data and safety monitoring board (DSMB) examined interim results to determine whether the trial should be stopped early, if the treatment proved either significantly beneficial or harmful to the experimental group relative to the control group (a practice known as a "stop rule").

By May 2002, an average of 5.2 years into the analysis (recall that women had entered the study over a five-year period), the DSMB found significantly higher risks of breast cancer, coronary heart disease, stroke, and pulmonary embolism in the experimental group, and that these increased probabilities outweighed some evidence of reduced risk of fractures and colon cancer. Therefore, the DSMB invoked the stop rule for the estrogen plus progestin component of the trial because it would be unethical to put more women at risk for adverse events by continuing HRT. Study results appeared in the *Journal of the American Medical Association*, and due to the results received wide publicity. A treatment that the medical profession had taken for granted as beneficial in reducing menopausal symptoms and preventing heart disease had faced a carefully designed randomized clinical trial. The results sent shock waves that continue to reverberate through the medical system years later. Further studies have repeatedly verified the original findings that women and their providers must balance any hormone replacement benefits against its demonstrable risks.

Sources: Writing Group for the Women's Health Initiative Investigators (2002); Fletcher and Colditz (2002).

Dummy Variables

Health care researchers often seek to find whether particular groups of patients or subjects differ from others. For example, Table 3.1 denoted men, African-Americans, and Hispanics, using 1 if the person was a member of such a group and zero otherwise. These groups were indicated by using binary, or *dummy*, variables. Econometricians may also wish to indicate whether research subjects are white (white = 1), or not (white = 0), or whether the subject is a woman (female = 1) or not (female = 0).

In Table 3.1 Regression (b) shows an important interaction between race and gender. In Regression (a), if we compare African-American men to white women, we add the coefficients 0.8967 (for male) + 1.6142 (for African-American) or find that on average the BMIs of the African-American men are about 2.50 points higher. Several government findings indicate, however, that many African-American women have high BMIs, due perhaps to diet or lifestyle preferences.[4] In Regression (b), we add an additional variable if the person is both African-American and male. It is statistically significant, and it gives an interesting finding. Relative to white women:

White men—BMI = + 1.2776
African-American women—BMI = + 2.5127
African-American Men—BMI = 1.2776 (men) + 2.5127 (Afr-Am) – 2.1862 (Interaction)
= + 1.6041

White men (+1.2776) and African-American men (+1.6041) have higher BMIs than white women. African-American women's BMIs (+2.5127) are close to another point higher.

Statistical Inference in the Sciences and Social Sciences

Natural scientists attempt to control experimentally, and not always successfully, for all of the other possible sorts of variation. By contrast, econometricians are seldom so fortunate. Experimental economic studies have been rare and expensive. One such study was the multimillion-dollar health insurance experiment conducted by the RAND Corporation in the late 1970s and early 1980s, funded by the federal government. We discuss parts of that study (Newhouse and collaborators, 1993) in several later sections of this text. Even with the careful planning that went into RAND's experimental design, the researchers could not avoid some major analytical issues.

Other fields have similar problems. A 1988 report from the Panel of the Institute of Mathematical Statistics referred to analytical problems in chemistry:

> The data are frequently complex with a large number of dimensions, may sometimes have a time element, and can be further complicated because of missing values. In some instances, standard multivariate or time-series methods may suffice for analysis, but, more commonly, novel developments are required, for example, to handle the problem of multivariate calibration.
>
> (Olkin and Sacks, 1988, pp. II–1)

Econometricians must most often use natural experiments and must seek ways to account for the other variations. Because many policies, such as the provision of public health services or the regulation of the prescription drug industry, depend on accurate measurement of economic phenomena, it is essential that the measurements be accomplished carefully and scientifically.

A 2011 report from the Center for Turbulence Research at Stanford University examines experimental uncertainties in determining chemical reaction rates, tracing them to problems in measuring activation energy and collision frequency. They note:

> very seldom can a reaction rate be measured in experiments without making use of another [sic] chemical reactions for its determination . . . This suggests that all uncertainties in chemical mechanisms are interwoven in a very experiment-specific manner, and that some of the regions predicted by independent random-variable models of

> uncertainties in multiple-reaction mechanisms are likely not physically accessible. Further research is needed to clarify these points.
>
> (Kseib, Urzay, and Iaccarino, 2011)

Conclusions

This chapter has provided a "taste" of the statistical methods necessary to address questions that occur in health economics and to clarify the analyses where statistical material is presented later in the text. Students should be able to formulate questions in terms of hypotheses, read statistical test results to determine significance of results, understand statistical significance, and interpret reported regression results. Emphasizing problems to watch for in statistical analysis does not mean that researchers should be unduly skeptical over the statistical data. On the contrary, we seek to distinguish the best studies in which to place our confidence.

Summary

1 Economists usually must collect information from people doing day-to-day activities and use statistical methods to control for the confounding differences among the people that they are analyzing. The more successful they are in controlling for such differences, the more reliable the analysis will be.
2 Statistical methods suggest formulating economic assertions as hypotheses, and collecting data to determine whether the hypotheses are correct.
3 Hypotheses that test for equality among two or more items are called *simple* hypotheses; hypotheses that test whether two or more items are greater (or less) than each other are called *composite* hypotheses.
4 To test hypotheses appropriately, the econometrician must:
   - state the hypothesis clearly,
   - choose a sample that is suitable to the task of testing,
   - calculate the appropriate measures of central tendency and dispersion, and
   - draw the appropriate inferences.
5 Regression analysis allows the econometrician to fit a straight line through a set of data points. In ordinary least squares regression, the sum of the squared deviations of the actual data points from the line is minimized.
6 R^2 measures the proportion of the total variation explained by the regression model. While it may be desirable to maximize R^2, most econometricians are at least as interested in the values of the estimated parameters.
7 Important skills in statistical analysis include:
   - understanding statistical significance,
   - interpreting reported statistical results, and
   - detecting problems in reported statistical findings.

Discussion Questions

1 List at least three ways in which natural experiments differ from laboratory experiments.
2 What is the difference between a simple hypothesis and a composite hypothesis? Why might economists choose one over another?

3 In considering the difference in smoking between men and women, what is the null hypothesis? What is the alternative hypothesis? Is the alternative hypothesis simple or composite?
4 Suppose that we wish to compare the health status of two groups of people. What variable might we use to measure the status? What variables might we wish to control in order to draw the appropriate inferences?
5 If someone reports that the mean weight for fourth-grade boys is 80 pounds and for fourth-grade girls is 78 pounds, what must you know to test hypotheses using the difference of means?
6 If we are trying to relate output to labor inputs and capital inputs using regression analysis, would we expect the coefficients of the regressions to be positive or negative? Why?
7 What are dummy variables? How are they useful in identifying differences among groups?
8 Suppose that you used regression methods to estimate the demand curve for physician visits and found a positive relationship: that is, you found that the higher the price is, the more visits are demanded. What problem has likely arisen? Explain the problem in words. Why might it make statistical inference difficult?
9 Rich people consume more health care services than poor people. Explain two ways one might test this hypothesis.

Exercises

(For students with access to statistical software or spreadsheet programs.) Consider the following data for a cross-section of individuals in the population, in which

Q = Quantity (in 100s) of aspirin purchased in a year
P = Average price of aspirin in that year
Y = Annual income
A = Age of buyer

Observation	*Q*	*P*	*Y*	*A*
1	1	1.5	20	25
2	2	1.5	40	20
3	4	1	12	25
4	2	1	10	30
5	2	1	8	30
6	3	2	30	35
7	3.5	1.5	30	40
8	4	2	20	40
9	7	1	20	45
10	1	3	15	40
11	2	2	18	30
12	3	2	20	32

continued

continued

Observation	*Q*	*P*	*Y*	*A*
13	3.5	2	15	36
14	4	2	10	30
15	2	3	25	20
16	1	4	15	25
17	8	2	15	55
18	9	1	40	50
19	1	4	10	45
20	10	1.5	30	55
21	6	1.5	35	60
22	2	1	30	40
23	3	1	25	40
24	3	2	20	35
25	3	2	15	35
26	4	3	20	35
27	1	4	20	25
28	1	4	25	30
29	2	5	28	30
30	3	1	30	32

Now consider questions 1 to 4:

1 If we divide the population into two groups, up to age 35 and over age 35, which group purchases more aspirin?
2 Divide the population into three groups—up to age 30, over 30 and up to 45, and over 45. Do the purchases vary by age?
3 What is the relationship in a regression analysis between Q and P? Between Q and Y? Between Q and A?
4 Calculate the multiple regression that relates Q with P, Y, and A. Which variables are statistically significant? What is the elasticity of Q with respect to P, to Y, and/or to A?
5 From Table 3.1, column b, suppose income is $20,000, the excise tax on cigarettes is $1, and the person is a 40-year-old white, non-Hispanic male who completed high school (education level = 9). Calculate the elasticities of demand for aspirin with respect to excise tax, income, and age.
6 Consider demand curves for aspirin, estimated for two different sets of consumers:
(a) $Q = 20 - 5P + 0.2Y$
(b) $Q = 30 - 5P + 0.2Y$
If Y = \$20 and P = \$1, calculate the price and income elasticities for group (a) and group (b). Whose elasticities will be higher? Why?

7 Given the regression estimate of the demand equation of

$$Q_X = 1{,}000 - 3.3P_x + 0.001Y$$

where Y is income, what is the change in demand if price rises by \$1, holding income constant? What is the percentage change in demand if price rises by \$1 from an initial price of P_x = \$200 given Y = 10,000? What is the effect on demand of a \$1 increase in income, holding price constant?

8 Consider the estimate demand equation of

$$\underset{}{Q_x} = \underset{}{1{,}000} - \underset{(3.5)}{3.3P_x} - \underset{(2.1)}{0.2P_z} + \underset{(0.5)}{0.001Y}$$

with t values in parentheses, where P_z is the price of another good Z, and Y is income. Is good Z a substitute or a complement? Can we say confidently whether good X is a normal good or an inferior good?

9 Look at Regression (b) in Table 3.1, and consider the following questions:
   (a) Does BMI increase as income rises?
   (b) Using the variables Age and Age-squared, calculate the changes in BMI between the ages of 25 and 35; 35 and 45; 45 and 55; 55 and 65. Are these changes constant? Why or why not?

10 Table 21.2 shows GDP/capita and total health care spending per capita for 42 countries (the first two columns of numbers).
   (a) Calculate the means of both variables.
   (b) Calculate a regression relating health care spending to GDP/capita.
   (c) Using the method discussed in equation (3.10) calculate the income elasticity of health expenditures.
   (d) What does your answer to part (c) indicate about the "share" of GDP/capita going to health? Why?

Notes

1 If we were working with a sample (as opposed to the entire population), we calculate the standard error by dividing by $n - 1$. All calculations are rounded to the nearest hundredth.

2 Although difference of means considers only two categories, *analysis of variance* methods allow the consideration of three or more categories in a more general way (difference of means is actually a restricted version of analysis of variance). Newbold, Carlson, and Thorne (2007) present good discussions on this and other statistical topics.

3 R^2 is often adjusted for number of explanatory variables k and number of observations n:

$$1 - \bar{R}^2 = \frac{n-1}{n-k-1}(1 - R^2)$$

Adding more variables k will always raise R^2 but may not raise $\bar{R}^2$, the "adjusted R^2."

4 See, for example, U.S. Dept of HHS, Office on Women's Health (2010), which indicates that about 4 in 5 African-American women are overweight or obese.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 4

Economic Efficiency and Cost-Benefit Analysis

In this chapter

- Economic Efficiency
- Cost-Benefit Analysis: Background
- Cost-Benefit Analysis: Basic Principles
- Measuring Benefits and Costs
- Valuing Human Life
- Cost-Effectiveness Analysis
- Cost-Utility Analysis, QALYs, and DALYs
- QALYs Revisited: Praise and Criticism
- Conclusions
- Appendix—Discounting

Health economics policy concerns require that we frequently and systematically evaluate alternatives. Just as rational individuals want to make the best choices given resource constraints, governments, too, face choices constrained by resource availability. For example, legislators and other policymakers must decide whether to spend more on preventive care versus giving more support to acute care facilities, or perhaps to medical research. When government regulates, the actual administrative expenditures may be relatively small. However, the economic consequences of the regulation can be very large, and evaluators must take corresponding care in evaluating the alternative scenarios. Economists base such decisions on the concept of efficiency. In developing the microeconomic tools for this textbook in Chapter 2, we explained the "welfare loss" caused by a monopoly's restricting quantity of production by charging too high a price. This welfare loss describes inefficiency—society has foregone opportunities for mutual gain. Efficiency applies to a broader range of phenomena than just monopoly, and we begin here by developing the concept more fully.

Economic Efficiency

Economic efficiency exists when the economy has squeezed out every opportunity for net benefits possible through voluntary means. Consider a single market, such as a local market for apples. Consumers' preferences for the apples can be measured by their willingness to pay for them; each person might have a different amount of money in mind. Likewise we measure the opportunity costs to society of an apple by the marginal cost of production. In some cases, there will be extra or "external" costs or benefits involved, but assume here that the private willingness to pay and the marginal costs summarize all benefits and costs. An efficient result for society will require a comparison of consumers' wants, as reflected in their demand, against the costs to society (either the private or public sectors) of the required production.

Economists use demand and supply analysis to define the efficient allocation of resources in competitive markets, and this idealized sort of market is convenient to explain the concepts. The competitive market is a market form that functions "properly." We will also see that for markets that do not function properly, or for cases where no markets exist at all, the underlying benefit and cost concepts often still apply.

The demand curve for apples represents consumers' willingness to pay for various amounts of apples. Marginal willingness to pay is another way of representing the demand function relating the quantity of apples demanded to the price of apples. Imagine, for example, lining up individual consumers from left to right in descending order on the willingness to pay for one apple. At small total quantities (along the market demand curve), the marginal willingness to pay is high, as only those who place considerable value on getting an apple are willing to pay for it. At larger quantities, the consumer's marginal willingness to pay is lower; additional consumers would not buy unless the price was lower.

At price P_1, in Figure 4.1, consumers together spend amount P_1Q_1. This is a fraction of what the apples are worth to the consumers, and it is also what they pay out. The total value of the apples to the consumers, however, also includes the additional shaded area under the demand curve, referred to as the consumers' surplus triangle. The consumers would have been willing to pay more than P_1, but did not have to do so. To see this, reconsider the example. Each consumer who buys an apple is willing to pay the price on the demand curve. That price reflects his or her benefit, but one must subtract the market price to get the net benefit. The shaded area then is simply the sum of the net benefits of each individual buyer.

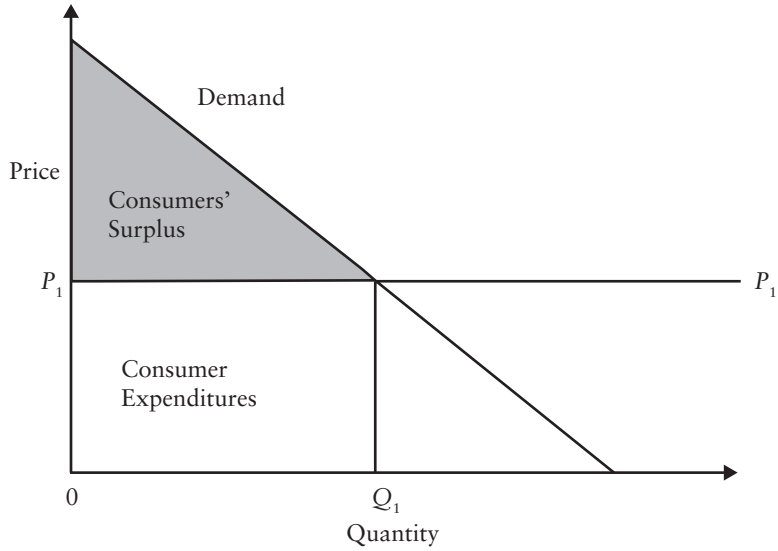


Figure 4.1 Consumers' Surplus

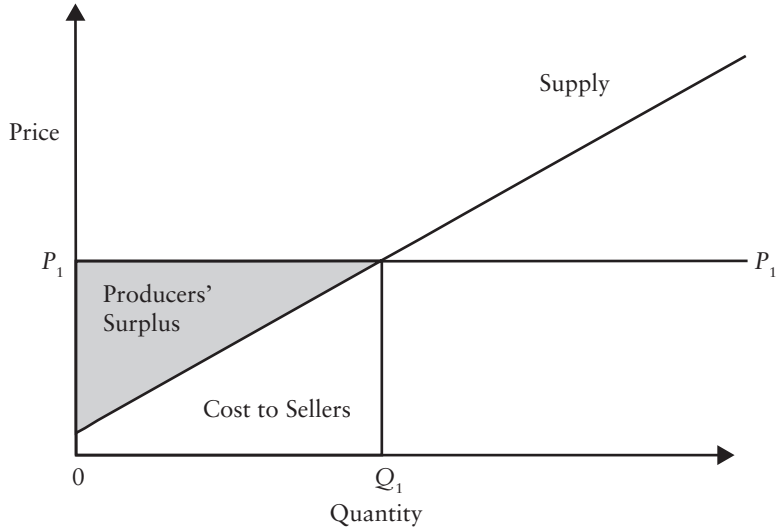


Figure 4.2 Producers' Surplus

Alternatively, the consumers' surplus at Q_1 equals the total value to consumers (the area under the demand curve up to Q_1) minus the amount that they must spend, P_1Q_1.

Figure 4.2 presents a supply curve which, with competitive markets, measures the marginal costs for producers to bring apples to market—the higher the market price offered to them, the higher the marginal cost they are willing to bear. Suppose that the market price of apples was $10 per bushel, but that some of the producers would supply apples even if the

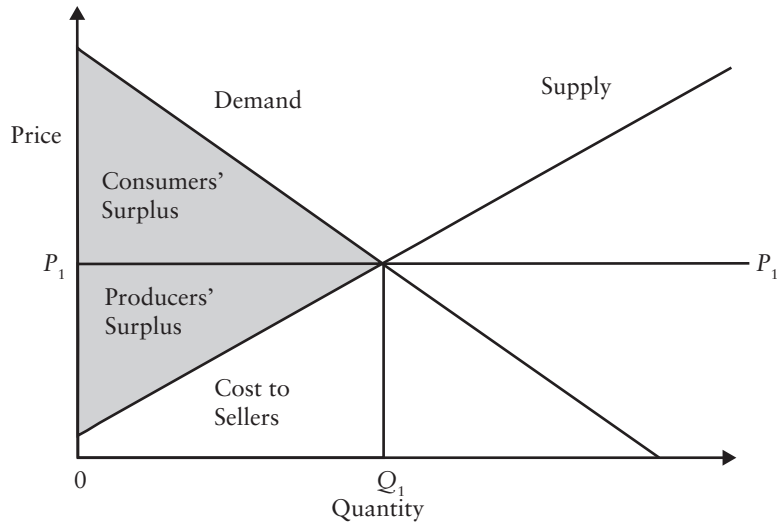


Figure 4.3 Efficient Quantity

price were only \$5. The \$10 per bushel that they receive constitutes a surplus of \$5 per bushel in excess of the price that is necessary to induce them to transfer resources from other uses to producing apples. Total consumer expenditures P_1Q_1 are apportioned into the resource cost to the sellers (the white area), plus the shaded area showing producers' surplus.

The economic criterion for maximizing well-being is to maximize the sum of the consumer and the producer surplus. Combining the supply and demand diagrams in Figure 4.3, we see the gains to consumers plus the gains to producers (the total net benefits to society) in the total shaded area.

Figure 4.3 shows that quantity Q_1, where demand equals supply, maximizes the sum of the surpluses. If quantity were less than Q_1, we could increase both consumers' and producers' surplus (the shaded areas) by increasing Q. If quantity exceeded Q_1, we could increase the surpluses by reducing Q because the incremental quantity (beyond Q_1) costs more (the supply curve) than it is worth to the consumers (the demand curve).

Efficiency requires that the optimal quantity come to market. In several places in this textbook we find that monopolies are economically inefficient by bringing too few goods to market. In contrast, polluters generally produce too many goods whose by-products such as smoke or untreated sewage impose costs on society that exceed the goods' market prices.

In competitive markets, supply and demand provide the efficient quantities of goods to the market—prices ration supply and demand according to consumer preferences and producer costs. However, students will recognize a wide range of goods for which such market signals are not readily available. These include bridges, parks, water purification systems, or mandated clean air. Decisions on whether to screen for breast or prostate cancers or whether to provide vaccines to the public, for example, depend on criteria that do not easily lend themselves to market tests.

With absent or incomplete markets, various evaluation tools have been developed to measure and compare project costs and benefits. Characterized in general as cost-benefit analysis, these tools seek to determine the appropriate quantity by measuring incremental or total costs, and incremental or total benefits. We address these methods in the following section.

Cost-Benefit Analysis: Background

Early forms of cost-benefit analysis (CBA) appeared over a century ago by the U.S. Army Corps of Engineers to evaluate flood control and other water systems. With the large public works projects during the Depression of the 1930s, it was necessary to justify expensive programs. The introduction of Medicare and Medicaid in 1965 led health economists to focus more on CBA. The federal government continued to improve and standardize methods, and since 1981 all new federal regulations must undergo CBA.

CBA measures the benefits and costs of projects in money terms. This often requires that we place dollar values on years of life or improvements in health and well-being. These challenges have led to the development of new ideas, and health analysts now use the general term *economic evaluation* to represent the entire collection of tools. Throughout the discussion, we will take the perspective of society as a whole rather than the narrower focus of the individual or firm. This is the appropriate perspective for public projects. We can similarly apply this logic to the investment decisions of a single hospital deciding whether to invest in a PET scanner, or even to the decision problem of a young man or woman considering a career in medicine.

Cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) have emerged as the principal alternatives to CBA. CEA applies to problems where policymakers accept the goal at the start and the problem is to find the least cost means to achieve it. CUA is a special form of CEA that introduces measures of benefits that reflect individuals' preferences over the health consequences of alternative programs that affect them.[1]

Measuring Benefits and Costs

Cost-benefit analysis addresses a wide range of measurement problems, often over many periods. It evaluates public investment costs, including those that have no markets to guide them. Likewise it investigates benefits that have no markets, to determine their prices. CBA also addresses public choices involving either benefits or costs that are external to the market they came from. Projects, such as putting a dam across a river or reconsidering the efficiency of a congressional program, are often controversial. Within health economics, controversies over the wisdom of immunization programs, patient screening, or heart transplants, for example, invoke many problems and criticisms that cost-benefit thinking must address.

CBA rests on the premise that a project or policy will improve social welfare if the benefits associated with it exceed the costs. These benefits and costs must include not only those directly attributed to the project but also any indirect benefits or costs through externalities or other third-party effects. Most simply, where B represents all the benefits and C represents all of these costs, a project is deemed worthwhile if $B - C > 0$. We can also rank projects according to the benefit to cost (B/C) ratio; thus, a higher B/C ratio generally indicates a project that will deliver greater social benefits for a given dollar of costs.

We measure all costs as opportunity costs, what we must give up to get what we want. The most common difference between public and private project evaluation is that public projects often have opportunity costs that have no market to serve as a guide for pricing. On the one hand, a dam project can destroy habitat for animal life, cover historical landmarks under water, and force whole towns to close down. On the other hand, the dam can enhance the fertility of otherwise barren land, provide recreational opportunities, attract waterfowl, and provide campsites and swimming areas for recreation. Many of these examples represent

either costs or benefits ordinarily bought or sold in markets. Many of the analytical problems stem from the imprecise task of placing dollar values on these difficult-to-evaluate costs and benefits. In some cases, there are methods for inferring the required values. For example, we can measure the benefits of a newly created lake view by observing how much the market values of nearby homes increase. For many cases, however, there are no easy answers.

The previous example also mentions a second problem. Public investments may produce costs or benefits to people who do not participate in them directly. Consider an immunization program that would offer protection against the flu to a wide cross-section of the population, concentrating in particular on the elderly and the ill. Analysts can measure the costs of materials and manpower directly, and estimate the benefits to those immunized. The program, however, also benefits people who never get immunized by reducing the number of infected carriers, thus reducing their exposure to the flu. These external benefits must appear in the CBA, but it may prove difficult to estimate them accurately.

A pollution clean-up program raises similar questions of how to treat externalities. For example, factories that discharge contaminants into the air or water create external costs by damaging the environment and adversely affecting third parties. Conversely, pollution abatement creates external benefits to others (e.g., boaters and home owners) who are not directly involved in the firm's decisions. They typically do not pay for the benefits they receive.

The analyses discussed assume that we have good measures of both marginal benefits and marginal costs. Consider, however, a project of size Q_1 as noted in Figure 4.4. Much of the research seeks a "value" for marginal benefit and/or marginal cost and applies it to all of those incurring the costs and/or the benefits. This is equivalent to using the

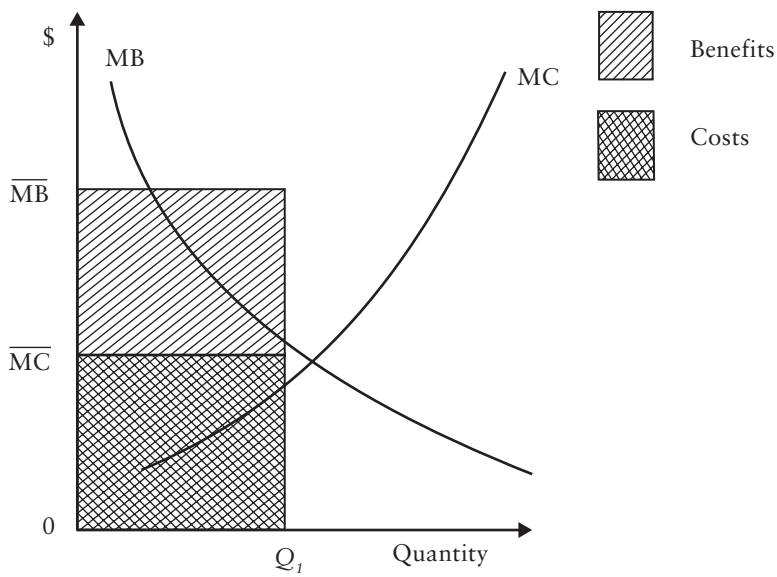


Figure 4.4 Gross v. "Micro" Measures in Evaluation

large rectangle for benefits, and the smaller one for costs, and then taking the difference (seeking B – C > 0), or the ratio (seeking B:C > 1). These measures of gross benefits and costs assume implicitly that all of the beneficiaries (and all of those bearing costs) are the same. Analysts have sought to determine marginal benefits in terms of willingness to pay or willingness to accept, but there has been little work on the cost side. Very often the data are incomplete.

In a series of influential articles, Ruger and colleagues develop a *micro-costing* methodology, a bottom-up approach to data collection that produces precise estimates by directly surveying the units and costs of each input consumed by participants in a given intervention. Micro-costing methods study interventions in particular settings, rather than in a very general group.

Ruger et al. (2012) demonstrate such a micro-costing model for addressing the costs of interventions to treat heroin addiction in Malaysia and 32 comparable countries worldwide. They seek to determine fixed costs, variable costs, and so-called societal costs (i.e., costs to individuals). Examples include:

Fixed Costs—Building and facilities, detoxification, training, quality assurance
Variable Costs—Therapy materials, testing materials, personnel costs, administration of medication
Societal Costs—Participants' time and travel costs, family members' time and travel costs, time in detoxification.

The researchers compare drug treatments naltrexone and buprenorphine.

They find that medication, and urine and blood testing, accounted for the greatest percentage of total costs for both naltrexone (29–53 percent) and buprenorphine (33–72 percent) interventions. In 13 countries, buprenorphine treatment could be provided for under $2,000 per patient. For all countries except the United Kingdom and Singapore, incremental costs of buprenorphine over naltrexone were less than $1,000. Looking at particular countries, they estimate that full treatment (100 percent) of opiate users in Cambodia and Lao People's Democratic Republic would cost $8 and $30 million, respectively.

Risk Equity versus Equality of Marginal Costs per Life Saved

Calamitous events often engage public sympathy to support rescue programs and emergency health care. These programs, in the abstract sense, seek to reduce health risks to the victims—for example, the risk of catching an infectious disease. Some argue that society ought to apply public resources so that health risks are shared equitably across the population. Perhaps equalizing life risks is impossible, but suppose that it were possible. Would it be the best choice for the use of society's resources?

Viscusi (2000) explains why such a plan would cause society to fall short of its welfare potential. Contrast a plan that follows sound economic principles. Were we instead to distribute public investments so that it equalized the marginal cost of a life saved across publicly funded programs, we would maximize lives saved for a given overall budget. The idea is simply that we should spend each next dollar where it does the most good (marginal analysis in CBA is discussed in detail later in this chapter). What about the interventions that we already have? Does U.S. spending follow this cost efficiency standard? Box 4.1 indicates that increased use of selected interventions could actually reduce net costs.

BOX 4.1

When Is Preventative Medicine a Good Investment?

When is preventive medicine a good investment? While some experts have suggested that clinical preventive services—such as immunizations, screenings, and counseling—are worthwhile when they save more money than they cost, others have suggested that a more appropriate standard should instead be that prevention offer good "value" for the net dollars spent. Maciosek and colleagues (2010) evaluate several readily available interventions including immunizations and screenings.

The study team examined the estimates of what costs and savings on the net impact on U.S. personal health care spending would have been in 2006 if a specific package of 20 evidence-based clinical preventive services had been used by 90 percent of the population for which each service was recommended. They calculated both the total costs and savings of providing the total package of services to 90 percent of the recommended U.S. population, and the additional—or marginal—costs and savings of increasing the use of the package from existing rates up to 90 percent.

The following table, derived from their work, ranks the top 7 interventions in terms of annual net medical costs per person by annual net medical costs per person per year. Negative values mean that money is actually saved!

Childhood immunizations, for example, would save 1,233.1 life-years per 10,000 children receiving them. The vaccinations costing $306 per year would actually save $573 in service, so the net saving *for the additional life-years* is $267 per person.

Life-Years Saved, Costs, and Savings with Negative Net Costs ($2006)

	Life-years saved per 10,000 people per year of intervention	*Medical cost of service per person per year*	*Medical savings of service per person per year*	*Annual net medical costs per person per year*
Childhood immunizations	1233.1	306	573	–267
Pneumococcal immunization	6.4	46	113	–67
Discuss daily aspirin use	63.0	21	87	–66
Smoking cessation advice and assistance	97.5	10	50	–40
Vision screening (adults)	2.1	5	22	–17
Alcohol screening and brief counseling	7.0	9	20	–11
Obesity screening	1.0	10	15	–5

The authors further calculate that increasing use of the package of 20 services from current levels to 90 percent costs less than the additional savings, resulting in a small negative net cost—or savings. The additional cost of increasing use from current levels to 90 percent would have been $18.3 billion, or 1.0 percent of U.S. personal health care spending in 2006. The savings resulting from increasing would have been $21.9 billion, and the net cost would have been –$3.7 billion, or –0.2 percent of U.S. personal health care spending in 2006, a net decrease!

Marginal Analysis in CBA

Figure 4.5 illustrates the marginal analysis principle applied to CBA. The marginal social benefits curve, representing the sum of all beneficial effects from increasing the abatement program by one unit, is downward-sloping. The marginal social costs, representing at each point the sum of all costs of increasing the program by one unit, is the *MSC* curve. For many pollutants, the marginal social benefits will include the benefits to the public of improved health. As an example, the incidence of cancer and respiratory disease has been linked to various forms of air pollution.

Society's maximum net benefit will occur where marginal social benefits equal marginal social costs. CBA represents an attempt to get the information with which to make the assessment. To illustrate the logic, a project requiring Q_3 is proposed, compared with the current discharge at Q_2. If *MSB* and *MSC* are properly measured, valid estimates of the project benefits equal the area under the *MSB* curve between Q_2 and Q_3, and similarly for costs by the area under the *MSC* curve. The net benefit equals area *A*.

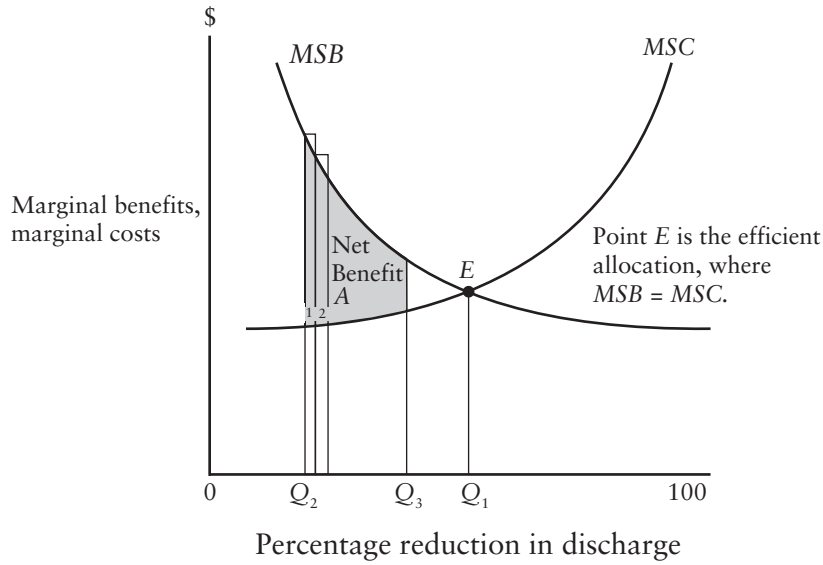


Figure 4.5 Efficient Use of Resources Where Marginal Benefits Equal Marginal Costs

Cost-benefit principles may help with such minor decisions as where to eat dinner, or what brand of lawnmower to buy. But the same principles can inform decisions that may very well affect the entire planet. The case at hand in Box 4.2 presents a modest invention that has intriguing possibilities beyond the village.

BOX 4.2

Cookstoves, Global Warming, Health in Developing Countries, and CBA

Much of the world burns biomass fuel for cooking and heating. Unfortunately this practice often results in smoky, unhealthful interiors, time lost in handling the material, and emissions of greenhouses gases. Cost-benefit analysis proves useful in identifying the best, or at least the most hopeful, solutions. A recent study by García-Frapolli et al. (2010) illustrates the method and some strong results. The study team focused on the Purépecha region of Mexico and estimated the benefits and costs per year of replacing peoples' current cookstoves with the more efficient Patsari design as follows:

Benefits/Year per cookstove	*$ (Benefits)*
Fuelwood savings	400.8
Job creation and income	19.1
Health impacts	208.6
Environmental impacts	103.2
Total benefits	731.7
Costs/Year per cookstove	*$ (Costs)*
Costs of cookstove construction	83.3
Indirect costs (dissemination)	25.3
Total costs	108.6
Ratio of Benefits to Cost*	9 to 11

Notes: *These ratios pertain to an estimated lifetime of the stove and they depend on the discount factor assumed. See the next section for more details on discounting.

For a given year, the benefit:cost ratio is almost 7:1, and because the improved stoves last longer, the lifetime ratio is between 9:1 and 11:1. With a benefit cost ratio this high, they concluded that the project was well worthwhile.

Discounting

Multi-period projects require that we discount the future costs and benefits to put them on an equal basis with present values. This necessity arises from one of two basic reasons which suggest that future dollars usually are not worth as much to people as present-day dollars.

First, a dollar today could be used for something else. That dollar could have purchased an interest-bearing instrument, such as a bond, which would have been worth the dollar plus the interest in next period's dollars. Conversely, next period's dollar must be discounted at the market rate of interest to discover its present value.

Second, many people tend to prefer the present when allocating spending. Which would you prefer, a dollar right now, or the same dollar ten years from now? Most would choose the dollar now. One consequence of this *time preference* is that the equilibrium interest rate will be positive, the reward for postponing consumption, and a cost for those who cannot wait or simply do not wish to. We discuss discounting in the appendix to this chapter.

Together, the time preference and the potential foregone interest from dollars spent on a project help explain why we discount future money values. The most commonly used method is relatively simple, although not without criticism (Frederick, Loewenstein, and O'Donoghue, 2002). Each period's costs and benefits are divided by a "discount factor" raised to the power t, the number of the time period, counting from the present as zero. The discount factor is $(1 + d)$, where d is the social discount rate, which for the present we assume is the market rate of interest. The present value equation is then

$$PV = \sum_{t=1}^{t=T} \frac{(B_t - C_t)}{(1+d)^t} \tag{4.1}$$

where t is zero for the initial period and increases up to the period T in which the project ends. For example, in the following period (usually taken as a year), the discount factor is raised to the power one, and the net of benefit over cost in this future period is divided by $(1 + d)$. Using a market interest rate of 6 percent for d, then the denominator becomes 1.06.

Notice also that if d is positive, the denominator in (4.1) becomes larger as t increases. This assumes that we discount the more distant future more heavily. This fact of discounting bothered both health care and environmental policymakers. Should those living in the present so disregard future generations? At first glance, time discounting may seem to guarantee that we will bequeath an unhealthful and polluted condition of life to the future generations. For reasons like these, some analysts propose that the social rate of discount, d, be set at a lower level than the current market interest rate or that we apply discount rate patterns that decline over time. The lower the chosen value of d (the closer to 0) the greater is the emphasis placed on the future.

Two ideas clarify economists' preference to use the market rate of interest. First, economists often temper the claims some people make on behalf of future generations with the knowledge that all human valuations must ultimately come from people living in the present. In a sense, we all speak for future generations through our actions in the capital markets.

Second, using market interest rates also has the advantage of measuring what people actually do as opposed to their responses to more hypothetical questions. Yet many would disagree. In any case, it is ultimately a social decision and one that has surprisingly large consequences. For a practical illustration of these consequences, see Box 4.3.

A number of other troublesome issues arise, however, when proposing the market interest rate for discounting public projects. Some find the approach inequitable when high and low income groups differ in time preferences. As the feature suggests, the more keenly felt problem at present is the prospective loss of human life and the possibility of permanent damage to the environment. The accompanying feature illustrates the large consequences of choosing a discount rate. Finally, some health economists propose to discount ordinary costs at the market rate but benefits to life and environment at a separate and lower rate (Brouwer et al., 2005; Claxton et al., 2006).

BOX 4.3

Discounting and Global Warming

In October 2006, the British Treasury issued the *Stern Review* (Stern, 2007), a disturbing report on the potentially cataclysmic consequences of greenhouse gases for our planet's future. It is widely credited with encouraging the world to understand the urgency of the problem. The discounting issues it raised help illuminate health economic discounting issues in general.

The *Stern Review* chose a discount rate of nearly zero to treat future losses as (essentially) equal to current losses. William Nordhaus (2007) criticized this choice, arguing that economies grow and that a zero discount rate ignores the fact that greater future wealth could make the estimated loss more affordable. When forced to choose, most economists recommend a discount rate of 3 percent, which is closer to average economic growth rates.

Opponents counter that the population also grows, and the gains from avoiding calamity will be spread over more people. Economists reply that a zero discount rate fails to reflect how real people actually treat future values versus present values, though many allow that environmental rescue ought to get a special rate, say, 1 or 2 percent. Should we discount the value of future lives saved at 0, 1, 2, 3, or 5 percent? Those who view this as a mere "academic" squabble may be the most surprised at the consequences.

Consider a calamitous climate event 100 years hence that would kill 1 million people, and measure the value of each life in the future as $6 million dollars. How much are we willing to invest today to save those future people? One hundred years from now the loss is $6,000 billion (or $6 trillion!), but what is the equivalent amount in today's dollars? The formula is

$$\text{Present Value} = \frac{\text{Benefit 100 years hence}}{(1+d)^{100}}$$

where d is the social rate of discount. The accompanying table reports those calculations. The rate chosen clearly makes a difference. With a zero discount rate, those future lives are valued at present at nearly one-third of the entire United States GDP. On the other extreme, with a 5 percent rate, those 1 million lives 100 years from now are viewed as worth less today than what Americans spend annually on "paper and allied products." (At a 5 percent rate $100, to be delivered 100 years from now, has a present discounted value of $0.76.) What rate d, then, is correct in the environmental context? The consequences of a discount rate choice are huge. The *Stern Review* has forced the right issues into vigorous discussion.

Benefit	*Social rate of discount, d*	*Net present value*
$6,000 billion benefit	0.00	$6,000 billion
100 years from now	0.01	$2,218 billion
	0.02	$828 billion
	0.03	$312 billion
	0.05	$46 billion

Risk Adjustment and CBA

To apply the market interest rate, however, we must recognize that there are many market interest rates. Riskier projects tend to have relatively higher rates of interest to reward investors, by the nature of the asset, are less sure of a full return. Often evaluators will adjust the social rate of discount to reflect the riskiness of the public project. Even this is conceptually not an easy task, however. Private markets for capital projects have private bidders and sellers, and they are often more capable of assessing perceived riskiness to themselves of adopting a prospective project. In contrast, public projects represent the public at large, whose view of the project's risk might be difficult to discern.

Nobel laureate Joseph Stiglitz (1988) has recommended that a relatively lower social discount rate better reflects the public's role. Stiglitz also proposes certainty equivalents to eliminate the biases that can result from including a risk-adjustment factor in the discount rate. Under the certainty equivalent method, the uncertain net benefit in any period, often represented by a probability distribution over project outcomes, is replaced by its equivalent (the value at which the decision makers are indifferent between the risky set of outcomes and a value received with certainty). The riskier a project, the lower will be the certainty equivalent to someone who is risk averse. The analyst then applies the cost-benefit criterion represented by equation (4.1) using the certainty equivalents for each period. The risk adjustment problem attracts the interest of health economists. Some researchers propose using a risk adjusted rate of return to compare programs with different risks (Sendi, Al, and Zimmerman, 2004).

Distributional Adjustments

Although cost-benefit analysis seeks primarily to improve efficiency, changes in the income distribution often result from a project. With narrowly focused projects, the tendency is to have a relatively small number of large gainers and perhaps many small losers. To the extent that society is concerned about equity, the distribution of the gainers and losers by income group should be a consideration. At the practical level, after ranking projects according to their net benefits, decision makers could invoke informal judgments as to the relative effects on the distribution of income and then adjust their rankings. More formally, Stiglitz proposes distributional weights through which the net benefits or losses give lower-income groups more weight than other groups. Of course, the method still will be subjective in that the weights themselves will necessarily reflect the judgments of the decision makers.

Inflation

Analysts also worry about inflation. Conceptually, macroeconomic inflation is not a problem. Because estimates of the inflation rate often turn out to be incorrect, it is best to measure both benefits and cost in current or real terms and then discount at the real (inflation-free) discount rate. If one introduces an inflation factor, then the discount rate should be increased by that inflation rate to get the nominal rate. It is important, though, that the discount rate reflect that inflation factor and not some other rate.

Valuing Human Life

A November 2013 report from the National Highway Traffic Safety Administration reported 33,561 traffic fatalities in 2012; this was 1,082 more fatalities than in 2011. A simple public health intervention would be the adoption of a 15 mile per hour national speed limit. No one

gets killed in collisions at this speed. While such an intervention would be costly in terms of travel time, wouldn't the lives saved be worth it?

Placing a value on human life is one of the most difficult but often unavoidable tasks in health care CBA. The first of several approaches, known as the human capital approach, estimates the present value of an individual's future earnings. This approach has been especially favored in legal applications that require estimates of damages. It also measures the loss of national output from mortality and morbidity or the production gains from saving and extending life.

In other ways, however, the human capital approach has flaws as a welfare measure. Retirees or crippled children, for example, have no value with respect to earnings. Also, the human capital approach does not directly measure people's willingness to pay to avoid risks of death, injury, or illness, nor does it measure what they are willing to accept as compensation for taking on such risks.

Willingness to Pay and Willingness to Accept

Consider the question, "How much are you willing to pay for the reduction in risks provided by new locks on your door?" The flip side of this concept is the compensation you would require to accept an additional risk to life and limb. The *willingness to accept* method also has many everyday examples. Its theoretical basis comes from the literature regarding the compensating differentials paid to laborers across various lines of work.

One of Adam Smith's successes in *The Wealth of Nations* was to explain why the wages of workers differed across jobs—explanations that still resonate well with modern economic theory. Smith's theory was the inspiration for modern-day economists to develop the theory of compensating differentials.

Contingent Valuation

It is often hard to derive a market test for willingness to pay for a risk-reducing medical treatment. Instead, the method of contingent valuation poses sets of medical contingencies such as: "*If* you faced an X% (high) risk of heart attack, how much would you be willing to pay for a medical procedure that would reduce your risk to Y%?" The set of questions, visual images (if used), and researcher interactions is called the format. Contingent valuation has enabled practical studies to move forward, but it has proved vulnerable to challenge. Some researchers view problems of "hypothetical bias," the idea that the subject will not or cannot answer hypothetical questions realistically. Changes in format can induce change in the valuations, a so-called framing bias (Whynes, Frew, and Walstenholm, 2005).

Potential resolution may come from studies that compare willingness to pay based on real-life behavior with measures of the same concept based on a question and answer format. Bryan and Jeott (2010) do this and conclude that the question-and-answer method does reasonably well for the given patient therapy they chose to study.

How Valuable Is the Last Year of Life?

Health system analysts commonly express puzzlement or dismay over "Why do we Americans spend so much of our health care dollar on the last year of life?" Later in this chapter we describe Quality-Adjusted Life-Years or QALYs and how this measure implies that a person has no claim on extraordinary health care when he or she gets very old. The elderly spend over one-quarter of their total health care expenditures on the last year of life. Even economists using standard utility analysis question why one's utility of life would get to be

so high when one has low natural prospects for living many years and when one's motive to bequeath to one's heirs is a ready alternative.

Becker, Murphy, and Philipson (2007) offer an interesting resolution to this puzzle. They begin by arguing against linearity in the calculation of the value of a life-year. For example, a commonly used value of $100,000 per life-year is based on a linear extrapolation for people's responses to marginal changes in their probability of death. It would seem irrational to us to save a life worth $100,000 by incurring expenses worth $500,000. But the $100,000 would be incorrectly applied to a patient's valuation of life when the risk change is life versus death *right now*. People facing a survive-or-die situation are not dealing with a marginal change: for example, a firefighter may increase his risk from 0.010 to 0.015, a marginal change. But the patient at life's possible end may be considering a hoped-for reduction in the probability of death from 1.00, if not treated, to 0.60, for example, if treated.

Their analysis rests on four intuitive ideas. First, for many of the very old and sick, their resources have very low opportunity costs because they cannot enjoy their wealth once they have died. Second, they may rationally have "hope" for living, including the hope that more advanced health care will be developed within their extended lifetime. Third, their "social" value of life (the value of their life not only to themselves but to family, friends, and community) may be very high. Finally, these authors show that the value of an extended life-year may be as high for frail patients as it is for those with better health.

Willingness to accept (WTA), derived from labor economic theory of compensating differentials, and willingness to pay (WTP), derived from consumer purchasing behavior for risk-reducing devices—are nearly two sides of the same coin. A basic difference, however, is that purchasing behavior (for WTP) is limited by a person's budget; we can afford to pay only so much to reduce our risks. In contrast, the WTA has no theoretical limit, a fact most relevant for the world's poor.

Researchers in health economics and other disciplines have applied both approaches. However, methods to elicit dollar values differ in practice. "Wage-based" estimates of willingness to accept risk for extra pay, for example, observe real-life behaviors. "Stated preference"

Table 4.1 How Much Is One Life Worth?

Meta-analysis of wage-based studies	*Years covered by the studies*	*Value of life in 2009 dollars*
Miller (2009)	1974–1990	5.2 million
Mrozek and Taylor (2002)	1974–1995	2.0 to 3.3 million
Viscusi and Aldy (2003)	1974–2000	6.9 to 9.5 million
Kochi et al. (2006)	1974–2002	11.1 million
Meta-analysis of stated preference studies	*Years covered by the studies*	*Value of life in 2009 dollars*
Kochi et al. (2006)	1988–2002	3.5 million
Dekker et al. (2011)	1983–2008	2.7 to 8.5 million
Lindhjem et al. (2010)	1973–2008	3.2 million

Note: These data are from Maureen Cropper, James K. Hammit, and Lisa A. Robinson (2011), "Valuing Mortality Risk Reductions: Progress and Challenges," Discussion Paper, Washington, DC: *Resources for the Future*.

methods ask subjects to state the risk/money trade-offs they would prefer. Table 4.1 illustrates the wide range of estimates from various applications. Many have studied the value of a statistical life (VSL), and the table summarizes them using meta-analysis. These estimates of VSL use simultaneously all the data from previous work and try to reconcile them to a single value or a range of values. Many choose to describe the value of a human life as a bit over $6 million dollars while $100,000 describes the value of one life-year (Cutler, 2007).

Researchers have sought to refine the VSL estimates. Aldy and Viscusi (2008) explore how and why VSL varies with age. On one hand, those aging have fewer life-years remaining, suggesting that VSL might decline with age. On the other hand, income and wealth grow with age, and then decline in the later years. It may not be possible to transfer wealth easily between age levels. The authors find that VSL increases in one's earlier years and then declines, following an inverted "U" shape over age.

Cost-Benefit Analyses of Heart Care Treatment

Notwithstanding the difficulties of CBA, one finds excellent examples in the literature. Examine a CBA by David Cutler (2007) that compared treatments for patients at risk for recurrence of heart attack. The treatment in focus is "revascularization," the use of bypass surgery and/or treatment with stents to improve blood flow to the heart. Cutler's study not only examines econometric issues but also estimates the patient's lifetime costs and benefits. To model the patient's lifetime, he acquired data to permit the study of 17 years into the future following the treatment. Table 4.2 reports the change in risk of death at each of the future years; it compares improvements to survival for patients admitted to a revascularization-capable hospital and those admitted to a High Volume (assumed to be high quality) hospital but one not having revascularization capability.

Table 4.2 Costs and Benefits of Medical Technology for a Lifetime

Time after MI	*Revascularization*		*High Volume Hospital*	
	Survival	*Spending ($)*	*Survival*	*Spending ($)*
1 Year	+0.061	30,149	–0.009	4,065
2 Years	–0.029	27,339	–0.005	5,300
3 Years	–0.067	25,919	–0.004	5,993
4 Years	–0.043	26,820	–0.001	6,560
5 Years	–0.106	27,517	–0.005	7,296
6 Years	–0.119	29,662	–0.005	7,659
7 Years	–0.119	31,090	–0.005	7,953
8 Years	–0.108	32,919	–0.004	7,982
9 Years	–0.111	36,961	–0.006	8,087
10 Years	–0.119	38,028	–0.007	8,314
11 Years	–0.113	38,191	–0.006	8,532
12 Years	–0.120	40,804	–0.009	9,002
13 Years	–0.074	38,079	–0.006	9,161

Table 4.2 *continued*

Time after MI	*Revascularization*		*High Volume Hospital*	
	Survival	*Spending ($)*	*Survival*	*Spending ($)*
14 Years	–0.064	38,708	–0.005	9,671
15 Years	–0.047	36,758	–0.005	9,524
16 Years	–0.041	37,200	–0.006	9,599
17 Years	–0.051	37,990	–0.007	9,770

Note: High Volume Hospitals are defined as hospitals that admit 75 or more heart attack patients in a year; these serve as comparisons and they are assumed to be high-quality hospitals in terms of heart care. High-quality hospitals do not necessarily have revascularization capability. "Survival" measures change in the risk of death (+ = higher probability) compared to the hospitals that have neither high volume (quality) nor revascularization capability.

Source: David Cutler, "The Lifetime Costs and Benefits of Medical Technology," *Journal of Health Economics*, 26 (2007): 1081–1100; data from his Table 5, p. 1094, with permission. Copyright © 2007 published by Elsevier B.V.

By analyzing all 17 years we see some unexpected patterns. Note that at Year 1, the revascularization estimate shows that it actually worsens the chances of survival. Cutler's columns labeled "Survival" are measured negatively as changes in the death probability, so a negative value indicates a reduction in death risk. The adverse Year 1 result quickly changes to improve the survival rate for many years. The cost of the advanced treatment is much higher than costs via the control group, those hospitals that lack the revascularization capability and yet treat heart cases in high volume.

Cutler calculated the increased life expectancy attributable to each of the two treatments. Revascularization increased life expectancy in this sample by 1.1 years (the sum of the revascularization survival rates) at a cost of approximately $38,000, thus achieving its gains at a rate of $33,246 for each life-year. The High Volume hospitals increased life expectancy by only 0.06 years, and even though their costs were low, their costs per life-year saved were $175,719. Estimating the value of a human life-year to be about $100,000, Cutler concluded that the $33,246 gain from revascularization easily proved cost beneficial.

Cost-Effectiveness Analysis

Given the difficulties of placing monetary values on life and health, as well as valuing other intangible benefits, cost-effectiveness analysis (CEA) sometimes provides a more practical approach to decision making than CBA (Garber and Phelps, 1997). CEA compares the costs of achieving a particular nonmonetary objective, such as lives saved. In cost-effectiveness analysis, one assumes that the objective is desirable even if the benefits have not been evaluated in monetary terms (strictly speaking, each project might yield negative net benefits were it feasible to compute those net benefits). Though the valuation of benefits is avoided, the problems of determining costs remain.

The proper evaluation of costs per output in CEA refers to the ratio of incremental costs to incremental output, as noted in equation (4.2). Let the change in social costs incurred due to

a particular project be $C_1 - C_0$, and let the gain in health output be $E_1 - E_0$. We then compare the various projects by the ratio:

$$\text{CEA ratio} = \frac{C_1 - C_0}{E_1 - E_0} \tag{4.2}$$

The costs as usual are in dollars, while the outputs are the chosen health status measure. To compare projects with this method, we must measure the outputs in the same units across all projects considered.

Advantages of CEA

The task under CEA is conceptually similar to a firm's production decision, which is to produce a chosen level of output at the lowest possible cost. Also, as in the firm-production decision, the objective must be quantifiable and measured in the same units across projects. Otherwise, a clear relationship between costs and output cannot be determined.

For example, the Department of Defense (where modern cost-effectiveness analysis began) has applied cost-effectiveness analysis to determine the most cost-efficient means of achieving a particular level of military preparedness. Quantifying objectives, analysts estimate the most efficient means of achieving the objectives. CBA is not feasible in such cases because of the difficulty in evaluating benefits of the military capability in monetary terms.

As another example, many communities support recycling to the point where mandatory recycling is becoming more widespread. Assuming that a community has decided on the goal of reducing garbage mass, it can use cost-effectiveness analysis to compare recycling with incineration and other waste-management strategies.

Finally, CEA can be a useful first step toward undertaking a cost-benefit study. If the analysts run into significant problems in undertaking a CEA, it is unlikely that a CBA will be feasible. Conversely, good progress in developing a CEA can often determine whether it is possible to take the next step and extend the CEA into a cost-benefit study.

Cost-Utility Analysis, QALYs, and DALYs

The pressing and frequent need for cost-effectiveness analyses of health projects might account for the development of practical variations of the technique of cost-utility analysis. Principal among these is the quality-adjusted life-year (QALY), which evaluates each project on the basis of its incremental costs per extra QALY delivered to the patients or other subjects (Garber and Phelps, 1997; Ried, 1998).

The QALY is a weighting system, typically designed by health professionals, that assigns a value, q, generally ranging from 1 (perfect health) to 0 (death) to represent quality of life for each year. In its welfare economic version, the weights for QALYs are in principle derived by eliciting the individuals' preferences for different states of health. Analysts often sum QALYs over groups of people. In this case, the procedure departs from standard welfare economics by its comparison and in some cases imposing of interpersonal "utilities" across people summed in a cardinal manner, practices that welfare economists generally avoid. QALYs of improvement are calculated as:

$$\text{QALY} = \sum_{t=1}^{t=\max} \frac{F_i q_i}{(1+d)^t} \tag{4.3}$$

where F_i is the probability that the person is still alive at age i; d is the time discount factor; and the value q_i is the quality weight, between 0 and 1, assigned to each year of the person's remaining life until a maximum value, *max* (Garber, 2000).

Suppose, for example, that a patient has the opportunity for a treatment that will extend life by one year with a probability of 0.9 (F_1 = 0.9) and by two years with a probability of 0.5 (F_2 = 0.5). The patient will die with certainty after two years. Quality weight q_1 is 0.8 in Year 1 and q_2 is 0.6 in Year 2. The discount rate is 0.05 per year. Thus, using equation (4.3) the QALY computation is:

$$\text{QALY} = (0.9 \times 0.8 \div 1.05) + (0.5 \times 0.6 \div 1.05^2) = 0.96$$

indicating that the expected effectiveness of the treatment is 0.96 QALYs. This serves as the denominator for equation (4.2). Costs per QALY can then be used to compare alternative interventions.

The production of QALYs provides a way to demonstrate the contrasts between equity and efficiency, and distinctions among utility maximization, benefit-cost, and cost-efficiency analyses. Figure 4.6 considers Ed and Harry, who each start with 10 QALYs at point *X*. If society gives its entire health budget to Harry (point *A*), it will achieve 10 additional QALYs. Harry will now have 20 in all; Ed still has 10. Likewise, if it gives the entire budget to Ed, it will achieve 20 additional QALYs or point *B* (we intentionally draw the graph to scale) for 30 QALYs in all. The production possibility frontier (PPF) between giving resources to Ed or to Harry is the curve connecting points *A* and *B*.

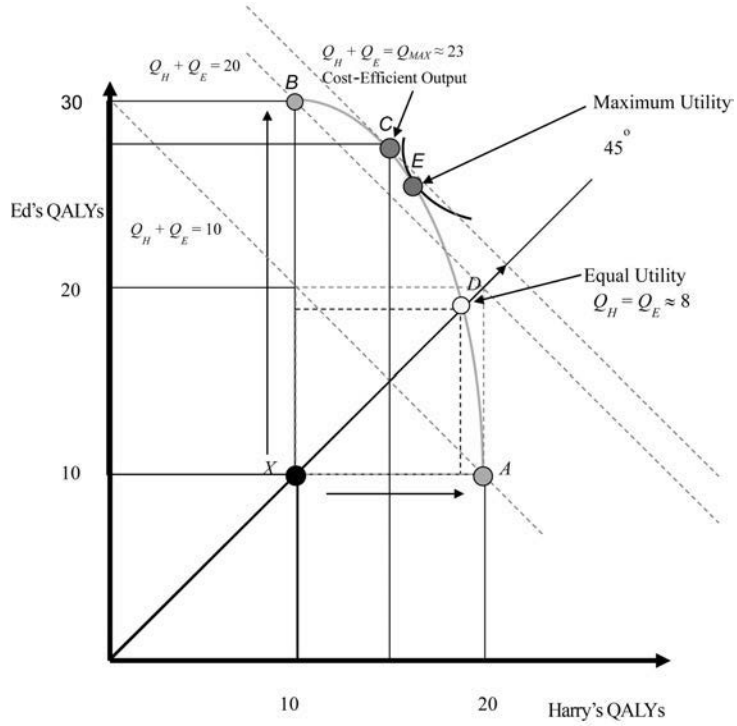


Figure 4.6 Cost Efficiency and Utility Maximization with QALYs

We can draw a set of lines with slopes –1 in the graph. Each line represents a constant number of QALYs. Clearly, we can produce more QALYs if we give all of the resources to Ed (20 vs. 10). However, if we draw yet another line (with slope –1) representing 20 QALYs from point *B*, we discover that we could produce still more QALYs in total. Although Ed is more productive totally, he is not more productive at the margin at point *B*. In short, Ed's twentieth QALY costs *much more than* 1 of Harry's foregone QALYs. Moving "southeast" down the production possibility frontier increases the total number produced. The maximum amount of QALYs occurs at point *C* where the slope of the PPF equals –1.0. This is the most "cost efficient" production point (the maximum production for a given amount of resources) for QALYs. Reading the graph shows a maximum of approximately 23 (additional) QALYs, with Ed having eighteen, and Harry having five.

Suppose we define the most equitable point as one where Ed and Harry have the same number of additional QALYs, or a 45-degree ray from the origin. Allocating resources "equitably" gives both Ed and Harry approximately eight QALYs at point *D*. Because Harry is relatively unproductive in making QALYs, assuring him the same amount as Ed directs relatively more resources to Harry, and reduces the *total* production of QALYs. Finally, a utility-maximizing allocation (at point *E*) may differ from all of the others.

The "take away" from this exercise is that different evaluation criteria lead to different results. QALYs eliminate the need for putting a dollar value on output such as life-years, while substituting an equally arbitrary metric (quality-adjusted life-years). Further, careful analysis shows that plausibly desirable outcomes (such as equal output for all) may come at very substantial resource costs.

QALYs Revisited: Praise and Criticism

Are QALYs Consistent with Standard Welfare Economics?

As we have discussed, cost-benefit analysis rests on the economic standard of efficiency, the Pareto Principle, which states that if an option of society improves the well-being of some people while harming no one then that option enhances welfare. Put simply, CBA has developed to be the standard of modern welfare economics. To say that CUA using QALYs is consistent with CBA would be to say that cost-utility analysis, too, chooses the efficient options for society; in the language of welfare economics, it would be "first best." Garber and Phelps (1997) and Bleichrodt and Quiggin (1999) argue that very restrictive underlying conditions would be required in theory to cause cost-utility analysis to attain the welfare economic standard. Blomqvist (2002) asserts that CUA cannot, as typically applied, attain the first best result just described.

Extra-Welfarism

Can a standard outside of standard welfare economics, one perhaps based on very different principles, give us a better world? Standards making this claim are called "extra-welfarist." Some extra-welfarist bases for decision making often discard the economist's idea of utility, and may reject the idea that people make their decisions rationally. Does it matter if we have two different approaches to health care project evaluation, one based within welfare economics and one outside of it?

Health economists Bala and Zarkin (2000) argue that consistency with welfare economics is important in evaluating a public project. The fundamentals of welfare economics

describe the net benefits to society of the project. Others argue that the willingness-to-pay principle is appropriately broader than extra-welfarist approaches. Besides counting patients' willingness to pay for extended life-years, it includes their willingness to pay for side benefits to recreation and family life made possible by the treatment (Olsen and Smith, 2001).

The extra-welfarists, however, point to inadequacies in the standard economic welfare framework. The welfarist view, they complain, commonly adds individual utilities without recognizing the interdependencies between people or their identification with the whole. Furthermore, the extra-welfarist approach using QALYs avoids a problem of willingness to pay; it avoids inequities that can be caused by an inequitable income distribution because ability to pay (due to higher income) is an important determinant of willingness to pay.

Amartya Sen (1985), a prominent critic of common conceptions of utility, proposes that each person is entitled to a life in which he or she can use a basic set of capabilities to achieve personal goals in life. Importantly, these capabilities would include basic health and functioning. Using this description, Cookson (2005) praises QALYs by showing that the quality index can be reinterpreted to represent a measure of Sen's capabilities.

What People Think

A developing criticism of CUA with QALYs focuses on the method's linear valuation of medical interventions as the simple sum of quality gains times life-years saved times the number of people treated. It doesn't matter who you are or what your situation is. A given improvement in functioning is valued the same regardless.

Dolan and colleagues (2005) tested these assumptions by reviewing literature sampling people about QALYs. Here are some examples of many differences they found in ordinary people's assumptions about QALYs.

- People are willing to sacrifice quality of life gains in order to give priority to the most severely ill.
- People dislike to discriminate by age, although they commonly weight the elderly somewhat lower.
- Health victims with dependent children are given more weight.
- People give much more weight to the health gains of people in the lowest social class.

The Ageism Critique

It has long been pointed out that QALYs tend to place a reduced value on older people when evaluating a medical intervention. A successful treatment of an old patient saves fewer life-years; those years are already limited by nature. Is this view of the elderly fair?

The issue goes deeper with the philosophical assertion of Williams (2001), who asked a stronger question: Is an extra life-year to be valued the same if an older person versus one younger receives it? His "fair innings" approach (coming from cricket, or in the U.S., baseball) argues that the younger person merits more concern (for U.S. readers, imagine a baseball game being ended after the third inning!), a normal human number of years. Note that in this version, the old person's shorter life expectancy is not the main issue; even an equal gain in QALYs is no longer assumed to have the same value between young and old.

An alternative approach, disability-adjusted life-years (DALYs), points out that we humans tend to be dependent on the middle-age groups when we are very young or very old. To the adherents of this view, the greater social-related weights should be placed on people in

the middle-age groups. The DALY theory provides a "hump-shaped" set of weights favoring the age groups in the middle.

How Are QALYs Used for Evaluation

Over the last three decades an uneasy consensus has developed that $50,000 per QALY represents an appropriate benchmark for medical interventions. Interventions that cost less than this would be looked upon favorably; those that are more expensive would require further justification.

Scott Grosse (2008) has examined this issue in detail. He terms the $50,000 threshold as an "arbitrary decision rule" that lacks theoretical or empirical justification. He continues that the World Health Organization cutoffs of one- and three-times per capita Gross Domestic Product have the advantages of international comparability and automatic indexing for inflation but lack a clear theoretical rationale.

He finishes his argument by noting that it is not clear that people's willingness to pay for health is closely related to QALYs or that QALY maximization describes preferences. Finally, he reflects an emerging consensus that higher dollar thresholds might be justified for life-saving interventions (such as organ transplants) than for interventions that reduce relatively mild symptoms.

Following up, Neumann, Cohen, and Weinstein (2014) agree with Grosse that the $50,000 threshold needs updating. They recommend that analysts use $50,000, $100,000, and $200,000 per QALY and argue that if one had to select a single threshold independent of the context of an explicit resource constraint or opportunity cost, they would suggest using either $100,000 or $150,000.

Conclusions

The evaluation of prospective health projects has generated substantial interest among health economists both in the theory and in the practical means to conduct evaluations to improve the society's well-being. Cost-benefit analysis requires the measuring of all benefits and costs attributable to the project both directly and indirectly. The need to identify external effects and to assign values to them in the absence of guidance from active markets poses both difficulty and controversy in practice. Future costs and benefits must also be adjusted to offset their differences from present values. Analysts debate discounting methods to accomplish this task in that projects whose benefits are achieved only in the more distant future tend to benefit from lower rates of social discount.

The most difficult task of all for CBA is probably that of attaching dollar values to human life. Cost-effectiveness offers a lesser but sometimes more practical evaluation result. By avoiding the assignment of dollar values to human life outcomes, the CEA focuses on providing useful guidance to the decision maker. The problems of measuring costs still apply, but it avoids the most arbitrary and controversial steps of valuing outcomes. The CEA approach requires that the outputs of the various projects be described in common terms. Several methods of output measurement and various discounting techniques have led to variations on CEA as a group called cost-utility analyses.

Most prominent among cost-utility analyses is to measure health output as quality-adjusted life-years, QALYs. These methods, and the debate over health economic evaluation in general, seek to improve health policy for society. At its simplest level, economic

evaluation recommends projects that achieve net positive benefits to society and prioritizes among these by their relative efficiency in doing so.

Despite its potential, CBA applications in health care still are less prevalent than one would expect. Difficulties in evaluating benefits, especially the value of life and improved quality of life, place limits on CBA and its usefulness to decision makers. As a result, cost-effectiveness analysis using QALYs has emerged as an important tool for program evaluation.

Summary

1 Economic efficiency requires the maximization of total welfare, with the optimum quantity reflecting the sum of consumers' and producers' surpluses.
2 Unlike private decisions made in the marketplace, (social) cost-benefit analysis (CBA) involves evaluation of social benefits and social costs in public project analysis. Often, markets do not exist to evaluate the benefits and costs of such projects.
3 CBA rests on the principle that society's welfare will be improved whenever the benefits of a project exceed its costs.
4 CBA represents an example of marginal analysis. The social optimum is achieved when marginal social benefit is equal to marginal social cost.
5 Though CBA appears simple, it can be difficult to apply. The difficulties include identifying all the relevant costs and benefits, including third-party effects, assigning monetary values, and making projections over many years for projects with long lives.
6 Micro-costing methods allow researchers to tailor their results to specific individuals, rather than imposing general measures. This allows for more precise calculation of cost-related evaluations.
7 The monetary values of future net benefits and costs must be discounted. Analysts debate the appropriate discount rate.
8 Health care projects must often value human life. The human capital approach and the willingness-to-pay approach have been the most widely applied methods.
9 Cost-effectiveness analysis can be used when it is difficult to place a monetary value on the benefits of a project. CEA is used to compare the costs under alternative projects of achieving some desired and quantifiable nonmonetary objective, such as the cost of detecting a case of cancer or the cost of a life-year saved.
10 Cost-utility analysis (CUA) is a special case of CEA in which the objective is measured in quality-adjusted life-years (QALYs) or some other indicator, such as disability-adjusted life-years (DALYs), that takes into account individuals' preferences for health.

Discussion Questions

1 Would the concept of consumers' surplus be a sound welfare measure if the income distribution were deemed inequitable? If so, in what way?
2 In what ways is social cost-benefit analysis similar to a consumer's decision about allocating resources or to a firm's investment decision? In what ways is it different?
3 What external benefits or costs would you expect from a project designed to develop sanitary waste product disposal in a third-world village? Why do these need to be considered as part of a CBA of the project?

4 Are jobs that are created as a result of a social project considered as a benefit, a cost, or both?
5 Some suggest that a dollar value cannot be placed on life; that is, life is priceless. Explain how the dilemma to social decision created by this view might be resolved.
6 Does it matter whether a higher or lower discount rate is applied to the CBA of a social project? If so, why?
7 Discuss possible reasons why the estimates of the value of life presented in Table 4.1 differ so much.
8 How does the willingness-to-pay principle of welfare economics differ from the valuation of an extra life-year in applying QALYs?
9 Figure 4.6 shows equity as point *D*, or equal outputs. If we defined equity in terms of equal inputs, how would our answer change? Why?
10 Distinguish between cost-benefit analysis (CBA) and cost-effectiveness analysis (CEA). Can CEA replace CBA in all cases? If not, why not?
11 The text noted that cost-effectiveness analysis began with the military. Consider the planning of a military intervention against an enemy. How (conceptually) would one measure benefits? Is it easy (or possible) to do so?
12 If a society has a fixed budget that it can devote to all interventions, formulate a prioritizing rule that would save the greatest number of years of life for a given budget.

Exercises

1 Using Figure 4.3, explain why a pollution abatement program that reduces discharge beyond Q_1 is inefficient.
2 Consider the following two projects. Both have costs of $5,000 in Year 1. Project 1 provides benefits of $2,000 in each of the first four years only. The second provides benefits of $2,000 for each of Years 6 to 10 only. Compute the net benefits using a discount rate of 6 percent. Repeat using a discount rate of 12 percent. What can you conclude from this exercise?
3 Consider the following table of costs and benefits from a governmental policy to clean the water in a local area.
   (a) What level of abatement is most efficient by general economic criteria?
   (b) Would a 70 percent level of abatement pass a cost-benefit test? Is it efficient?
   (c) How would you respond to those who argue for 100 percent abatement?

Level of Abatement (%)	*Total Costs ($)*	*Total Benefits ($)*
0	0	0
10	10	80
20	22	150
30	40	200
40	70	240
50	105	280

continued

Level of Abatement (%)	*Total Costs ($)*	*Total Benefits ($)*
60	150	320
70	210	350
80	280	375
90	350	385
100	420	390

4 Consider a project that costs $10,000 today. It will provide benefits of $4,000 at the end of Year 1, $3,500 at the end of Year 2, and $3,500 at the end of Year 3. If the discount rate is 6 percent, will this project be approved using cost-benefit analysis? Would your answer change if the discount rate is:
   (a) 5 percent?
   (b) 4 percent?

5 Consider a hypothetical three-stage screening test for a cancer with the following rates of detection and costs:

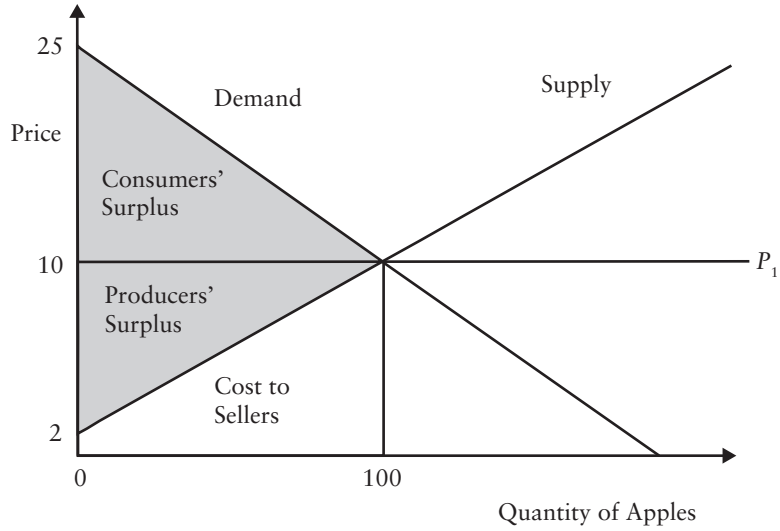
Stage	*Number of Cases Detected*	*Total Costs*
1	100	$200,000
2	105	260,000
3	106	300,000

   (a) Calculate the average cost per cancer detected in the three stages.
   (b) Calculate the marginal cost per cancer detected in the three stages.
   (c) Suppose that the marginal benefit per treated case is $12,000 per person. What would be the optimal screening, given the costs?

6 Using the diagram below, calculate:
   (a) total consumer expenditures
   (b) total cost to sellers
   (c) total consumers' surplus
   (d) total producers' surplus
   (e) the sum of the consumers' and the producers' surplus.

7 Using the diagram below, suppose that producers need to have licenses to sell apples, and that only 90 units of apples are licensed (i.e., Q is limited to 90). Calculate:
   (a) the sum of the consumer surplus and producer surplus
   (b) the reduction in consumer well-being because of the licensing.

8 In Figure 4.6 suppose that the production possibility frontier was a straight line with slope –2.0. What would be the cost-effective allocation of resources? Why is this the case?



Appendix—Discounting

Economists use discounting to compare streams of returns and/or costs over a number of periods. An analyst might be asked to compare Investment A, which provides $2,000 at the end of Year 1 and $2,000 at the end of Year 2, with Investment B, which provides $1,200 at the end of Year 1 and $2,900 at the end of Year 2. Although Investment B returns $4,100 over the two years compared to $4,000 for Investment A, most of the return on Investment B comes later, at the end of Year 2. It is often crucial to compare the investments with a criterion that considers the timing of the returns. *Discounting* is the analytical tool that analysts use for such comparisons.

Suppose George is offered the opportunity to buy a bond that will return $1,000 one year from now. How much is he willing to pay now? George always has the option to keep his money and earn interest rate r. He will buy the bond if he can pay a price far enough below the $1,000 return next year such that that price, multiplied by one plus the interest rate, equals $1,000. Algebraically, if the rate of interest is r and the unknown amount is x_1, then:

$$x_1(1+r) = 1{,}000$$

The value for x_1, then, can be solved as:

$$x_1 = \frac{1{,}000}{(1+r)}$$

If the rate of interest is 5 percent (or 0.05), then x_1, the discounted value of $1, one year hence, equals 1,000/1.05, or $9,524.

Similarly, the discounted value of $1, two years hence, is:

$$x_2(1+r)^2 = 1{,}000$$

The value for x_2, then, can be solved as:

$$x_2 = \frac{1{,}000}{(1+r)^2}$$

If, again, the interest rate is 5 percent, then x_2 equals $1{,}000/1.05^2$, or 90.70 cents.

Returning to the preceding example, we can calculate that the present value (the sum of x_1 and x_2), or *PV*, for Investment A will equal \$3,719, or $(2{,}000/1.05 + 2{,}000/1.05^2)$. The *PV* of Investment B is \$3,773. Thus, George will prefer Investment B.

In summation notation, where term $\sum$ represents the summation from period 1 to period *T*, the present value of a stream of returns *R* and costs *C*, over time, is:

$$PV = \sum_{t=1}^{t=T} \frac{(R_t - C_1)}{(1+r)^t}$$

It is easy to demonstrate that the relative *PV*s of Investments A and B may depend on the interest rate (or in evaluation studies, sometimes referred to as discount rate *d*) chosen for the analysis. If an interest rate of 15 percent were used instead of 5 percent in the previous example, the *PV* for Investment A would be \$3,251, compared to \$3,236 for Investment B. Intuitively, the higher interest rate gives the larger *but later* return in Investment B less weight.

Note

1 Gold et al. (1996) and Drummond et al. (1997) provide excellent comprehensive analyses of many of the issues raised in this chapter. Clifford Goodman (1998) provides a concise summary of the literature as well as a practical guide to the evaluation process.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 5

Production of Health

In this chapter

- The Production Function of Health
- The Historical Role of Medicine and Health Care
- The Production Function of Health in the Modern Day
- Do Other Measures of Health Care Affect Health?
- The Role of Schooling
- Conclusions

The production of health presents a central concern to the health economist and to public policy. Consider that the role of health care in society, including medical care provided by physicians, is ultimately a production question. We must learn about the determinants of health and about the contribution of health care. We can then better understand what decisions, both personal and public, will best produce health.

In medical terminology, this chapter addresses the efficacy and effectiveness of all those features of life, not only medical care, that plausibly contribute to our health. Unlike the typical doctor in practice, however, we look for evidence of the response of a "treatment" in the change in the health status of populations, as opposed to the treatment response of a medicine for the individual patient. We will see that the two approaches must remain in harmony and that both are fundamentally searches for causal relationships.

The Production Function of Health

A production function summarizes the relationship between inputs and outputs. The study of the production of health function requires that we inquire about the relationship between health inputs and health. The answers that economists and medical historians offer to this question surprise many people. First, the contribution of practitioner-provided health care to the historical downward trends in population mortality rates was probably negligible at least until well into the twentieth century. Second, while the total contribution of health care is substantial in the modern day, its marginal contribution in some cases is small.

This distinction between total and marginal contributions is crucial to understanding these issues. To illustrate this distinction, consider Figure 5.1A, which exhibits a theoretical health status production function for the population. Set aside the difficulties of measuring health status in populations, and assume that we have defined an adequate health status (*HS*) measure. Health status here is an increasing function of health care. Also, to avoid a perspective that is too narrowly focused on health care, we specify further that health status depends at least upon the population's biological endowment, environment, and lifestyle.[1] Thus, $HS = HS$ (Health Care, Lifestyle, Environment, Human Biology). Improvements in any of these latter three factors will shift the curve upward.

A production function describes the relationship of flows of inputs and flows of outputs over a specified time period, so the inputs and output in Figure 5.1A are measured over an implied period, such as a year. In practice, we might use the number of healthy days experienced by the population per capita, mortality rates, or disability days, to indicate health status.

To simplify the depiction, we have reduced all health care inputs into one scale called Health Care. In reality, health care consists of many health care inputs. Some of them include medical care provided by doctors of medicine or osteopathy, but other health care professionals also provide care. Conceptually, the health care measure, *HC*, may be thought of as an aggregate of all these types of health care, the aggregation being based on dollar values.

The marginal contribution of health care is its marginal product, meaning the increment to health caused by one extra unit of Health Care, holding all other inputs constant. Increasing Health Care from zero to one unit in Figure 5.1A improves health status by ΔHS_1, the first unit's marginal product. Numerically, this first unit of Health Care has increased the health status index from 32 to 43; $\Delta HS_1 = 11$ Health Status units. The next unit of medical care delivers a marginal product of $\Delta HS_2 = 7$, and so on.

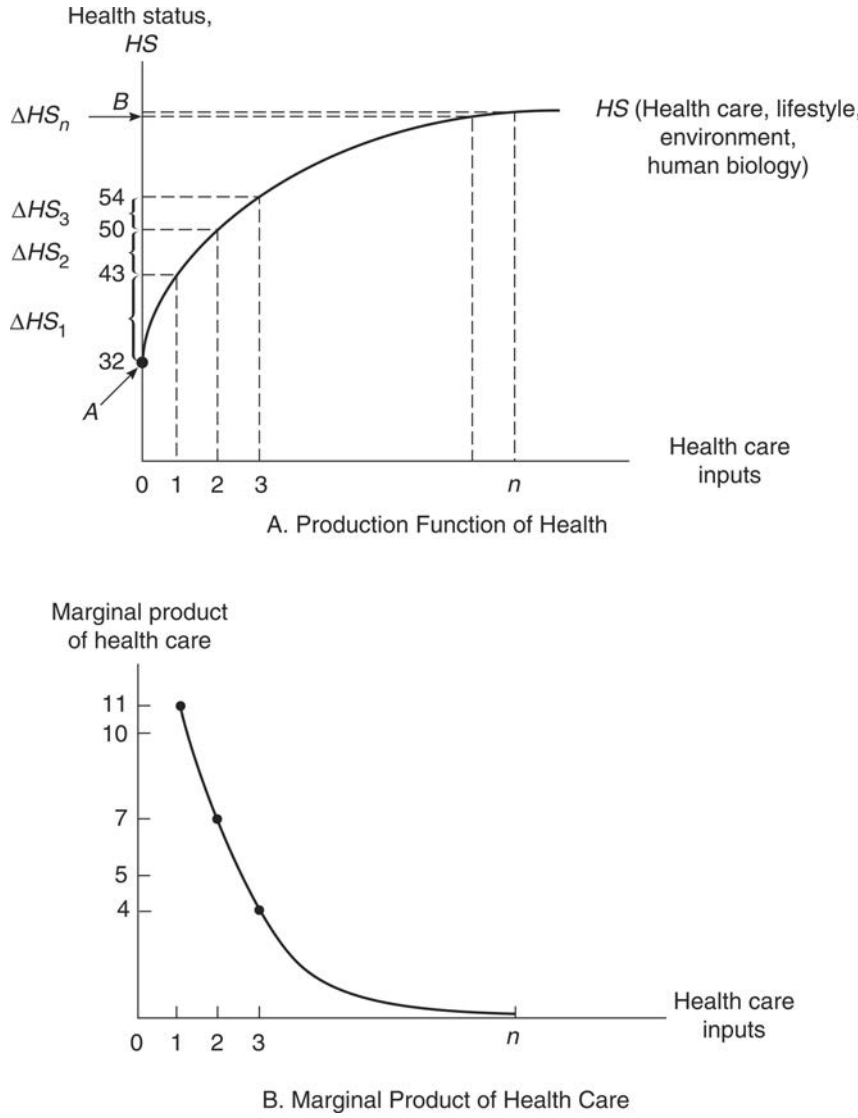


Figure 5.1 Production of Health

These marginal products are diminishing in size, illustrating the law of diminishing marginal returns. If society employs a total of *n* units of Health Care, then the total contribution of Health Care is the sum of the marginal products of each of the *n* units. This total contribution as shown, *AB*, may be substantial. However, the marginal product of the *n*th unit of medical care is ΔHS_n, and it is small. In fact, we are nearly on the "flat of the curve." Marginal product is graphed on Figure 5.1B.

We have drawn the health production function as a rising curve that flattens out at higher levels of health care but never bends downward. Would the health production function eventually bend downward? Is it possible to get too much health care so that the health of the

population is harmed? This is a logical possibility under at least two scenarios. *Iatrogenic* (meaning provider-caused) disease is an inevitable by-product of many medical interventions. For example, each surgery has its risks. Combinations of drugs may have unforeseen and adverse interactions. If the rate of iatrogenic disease does not fall while diminishing returns sets in, it is possible for the balance of help and harm from health care to be a net harm.

Medical scientists, such as Cochrane (1972), have pressed the case that much medical care as often practiced has only weak scientific basis, making iatrogenesis a real probability. Writing for the public audience, Dubos (1960) and Illich (1976) once warned of a medical "nemesis" taking away our abilities to face the natural hardships of life by "medicalizing" these problems. Illich argued that this medicalization would lead to less personal effort to preserve health and less personal determination to persevere; the result becomes a decline in the health of the population and thus a negative marginal product for medical care.[2]

Return to the distinction between total product and marginal product. Often, the marginals, rather than the totals, are relevant to policy propositions. For example, no one seriously recommends that society eliminate all health care spending. However, it is reasonable to ask whether society would be better off if it could reduce health care expenditures by $1 billion and invest those funds in another productive use, such as housing, education, transportation, defense, or other consumption. We could even reasonably ask if health itself could be improved by transferring the marginal $1 billion to environmental or lifestyle improvements.

Many of our government programs encourage health care use in certain population groups, such as the poor and elderly. Other programs, such as tax preferences for health insurance, provide benefits for those who are neither poor nor elderly and encourage their consumption of health care. The theoretical issues raised here suggest that we question the wisdom of each of our programs. The theoretical questions can be investigated with data of several kinds either directly or indirectly relevant to the production of health issue. We begin with the historical role of medicine, which indirectly bears on the issue of health production. After providing an overview of these efforts, largely the work of medical and economic historians, we then turn to econometric studies of the modern-day production function.

The Historical Role of Medicine and Health Care

Many medical historians agree that practitioner-provided medical interventions played only a small, perhaps negligible, role in the historical decline in population mortality rates. Effective medicine is a fairly recent phenomenon, and the delivery of effective medical interventions on a scale sufficient to affect population health indicators most likely appeared only well into the twentieth century. Though the magnitudes of other causes of mortality declines are still disputed, it is clear that a larger role, one of the most significant ones, might be attributed to public health measures and the spread of knowledge of the sources of disease. However, a number of scholars in this field attribute the largest share of the credit to improvements in environment, particularly to the greatly increased supply of foodstuffs that became available due to the agricultural and industrial revolutions.

The Rising Population and the Role of Medicine

The notion that medicine played a relatively minor historical role is certainly not new, and it has been asserted by researchers of various ideologies. This point of view is associated with the work of Thomas McKeown (1976), who focused on the dramatic rise in population in England and Wales from 1750 to the modern day.

The pattern of world population growth, including population growth in England and Wales, has interested many scholars, including McKeown. World population is hard to estimate for the distant past, but research by the United Nations (1996) and others shows that something extraordinary happened during the last 300 years. In the first century the population was roughly 300 million. For a thousand years thereafter, until the era of Viking ships, little or no change occurred. By the Age of Enlightenment, starting just before 1700, the population may have risen to 600 million. Then things began to change rapidly. Within a single century, the world population passed 1 billion people. The next 5 billion arrived within a mere 200 years. What had happened? Figure 5.2, based on United Nations data, reveals this startling pattern.

Returning to the history of England and Wales, the large rise in their populations in the period following 1750 is to a large degree a story of the population's health. Population increase comes from increased birth rates, reduced mortality, or increased net in-migration. Migration was not an important source of population increase in England and Wales; when accurate birth rate and death rate data became available from 1841, these data alone proved able to account for the population change. Likewise, fertility probably did not account for the change because recorded birth rates have declined during the period since data have become available. Declines in birth rates are a common finding in countries undergoing industrialization and modernization. In contrast, recorded mortality rates did decline substantially.

McKeown began by investigating which diseases contributed to the decline in death rates. Mortality data are very limited prior to the mid-1800s, but the records revealed an emerging picture. Table 5.1 shows death rates by disease category for three time periods. The table

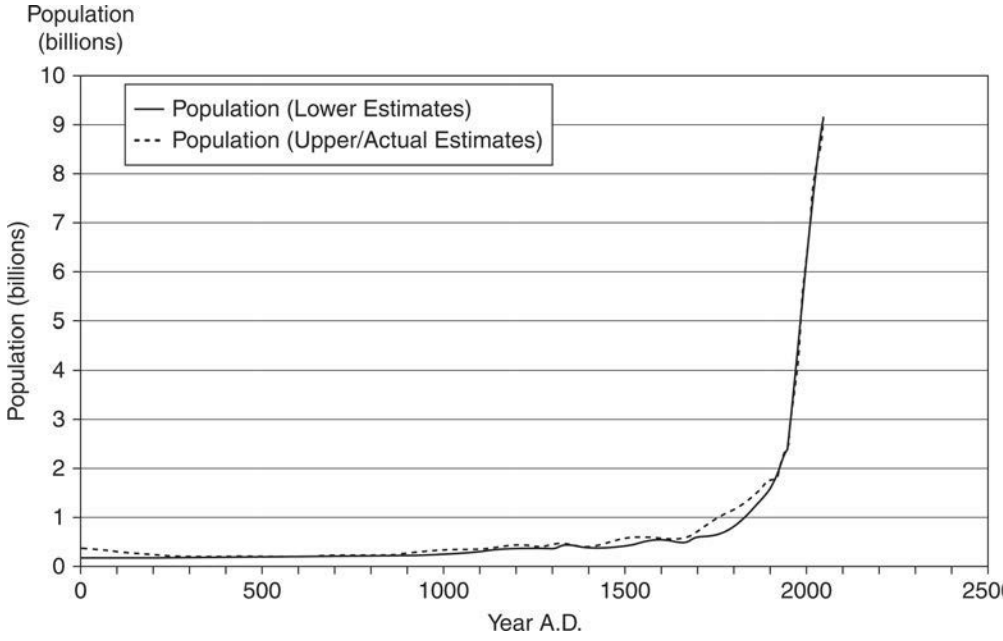


Figure 5.2 World Population from 1 C.E. to the Modern Day

Source: www.ciese.org/curriculum/popgrowthproj/worldpop.html, accessed November 2016. Data from U.S. Census Bureau.

Table 5.1 Death Rates* (per million) in 1848–1854, 1901, and 1971

	1848–1854	*1901*	*1971*	*Percentage of Reduction (1848–1854 to 1971) Attributable to Each Category*	*For Each Category, Percentage of Reduction (1848–1854 to 1971) That Occurred before 1901*
I. Conditions attributable to microorganisms:					
1. Airborne diseases	7,259	5,122	619	39	32
2. Water- and foodborne diseases	3,562	1,931	35	21	46
3. Other conditions	2,144	1,415	60	12	35
Total	12,965	8,468	714	72	37
II. Conditions not attributable to microorganisms	8,891	8,490	4,070	28	8
All diseases	21,856	16,958	5,384	100	29

Note: *Standardized to the age/sex distribution of the 1901 population.
Source: Reprinted from McKeown, Thomas, *The Modern Rise of Population*. New York: Academic Press, 1976 (p. 54).

shows that airborne infectious diseases account for the largest single portion of mortality reduction, and waterborne infectious diseases also make up a substantial portion of known causes. Regarding the airborne diseases, other data suggest that the main airborne diseases showing a decline in mortality include tuberculosis, bronchitis, pneumonia, and influenza.

What Caused the Mortality Rate Declines? Was It Medicine?

Many presume that the declines in the mortality rates were due to improvements in medical science provided to the public through medical practice, but counterarguments to this proposition bring it into question. In most cases, an effective specific medical intervention was not available until late in the period, well after the greater part of the mortality decline had occurred.

The argument can be illustrated for the cases of respiratory tuberculosis and a group of three upper respiratory diseases—bronchitis, pneumonia, and influenza. Mortality rates for these diseases fell to relatively low levels prior to the availability of effective medical interventions, whose availability occurred respectively after 1930, and for some cases well into the 1950s and 1960s. The picture is shared by waterborne diseases. About 95 percent of the

mortality declines in cholera, diarrhea, and dysentery occurred prior to the 1930s, when intravenous therapies became available. Likewise, typhoid and typhus mortality already had fallen to low levels by the beginning of the twentieth century. The pattern McKeown found for England and Wales also can be illustrated for the United States. McKinlay and McKinlay (1977) provided data for the United States from 1900 to 1973. Figure 5.3 shows these patterns for several infectious diseases. In most cases, as is shown, the availability of the effective medical intervention occurs well after the majority of the mortality declines.

BOX 5.1

Tuberculosis and *The Magic Mountain*

Thanks to the efforts of writers such as Nobelist Thomas Mann, the tuberculosis (TB) sanitarium of 100 years ago has found a permanent place in literature. Mann's novel *The Magic Mountain* describes with a mastery of medical detail, often admired by physicians, the characters' struggle with the deadly disease. Mann's incomparable development of character and dialog bring a reality to the society unique to this sanatorium, representing one he had himself visited. Hans Castorp, a young German engineer, is smitten with Clavida Chauchat, a young woman at the "good" Russian table. The Italian Settembrini's intellectual arguments with Naptha can result only in a duel. We grieve when Hans's cousin Joachim, a good, simple military man, succumbs to TB and his body is sledded down the mountain. Hans survives to return to the lands below.

Though effective chemical interventions were not available until after 1940, doctors did treat TB prior to 1940, notably with the widespread use of sanatoria such as the one in *The Magic Mountain*. Declines in TB mortality during the period studied by McKeown represented perhaps the most important example of declines in the mortality rate. Can we credit the sanatoria for this progress? This is unlikely, because it is unlikely that their capacity was ever large enough to affect the pattern of mortality in populations. For an account of the retreat of tuberculosis, see Smith (1988).

One of the most important changes in mortality in the twentieth century was the decline in infant mortality. Does this type of mortality follow the same pattern? A highly readable account of the modern historical pattern of infant mortality is offered in Victor Fuchs's *Who Shall Live?* (1975). Fuchs noted that infant mortality rates in New York City improved markedly from 1900 to 1930 and that this decline was due to declines in deaths from "pneumonia-diarrhea" complex. Fuchs concluded: "It is important to realize that medical care played almost no role in this decline. While we do not know the precise causes, it is believed that rising living standards, the spread of literacy and education, and a substantial fall in the birth rate all played a part" (p. 32).

Antimicrobial drugs were introduced in the 1930s. Between 1935 and 1950, the fall in infant death rates accelerated. Fuchs proposed that medical advances and rising living standards both contributed to the reduction in infant deaths during this period. Declines in infant deaths flattened somewhat beginning about 1950 but resumed a stronger decline about 1965. But the specific effective curative medicines of the twentieth century cannot explain the historical mortality declines. Were there other tools in the physician's black bag that were effective before 1900? Unfortunately, this too is unlikely. The problem is that there probably were

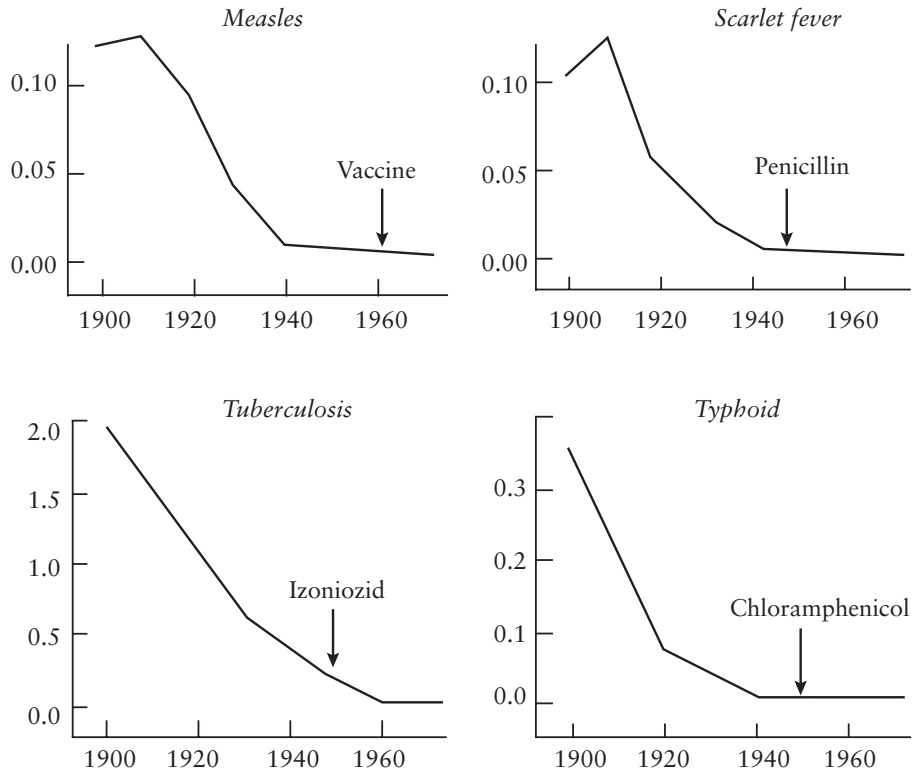


Figure 5.3 Fall in the Standardized Death Rate per 1,000 Population for Four Common Infectious Diseases in Relation to Specific Medical Measures for the United States

Source: Reprinted from *Milbank Memorial Fund Quarterly/Health and Society*, John B. Mckinlay and Sonja M. Mckinlay, "The Questionable Contribution of Medical Measures to the Decline of Mortality in the United States in the Twentieth Century," *Milbank Memorial Fund Quarterly/Health and Society* 55 (1977): 405–428, with the permission of Blackwell Publishers.

few effective tools available until well into the twentieth century. Even a clear knowledge of what caused disease was not widespread until 1900.

NUTRITION REDUCED MORTALITY Two of the most respected students of the mortality decline, medical historian Thomas McKeown (1976) and economic historian Robert Fogel (2004), argued strongly that the main cause was improved nutrition. McKeown reasoned by process of elimination. As we have just seen, he showed that medical interventions could not have been the cause, a claim that is still widely accepted. He considered other possibilities one by one. For example, some have suggested that perhaps the infectious organisms had spontaneously mutated and become harmless; he pointed out that the chances were remote that so many independent organisms had randomly mutated at about the same time.

McKeown also dismissed public health as a major cause, however, and this argument was to become controversial. If we re-examine his work in Table 5.1, we see that the largest

portion of mortality decline from 1848 to 1971 was due to declines in mortality from airborne diseases. He argued that public health projects, which focused on improving water quality and the safety of food, could have little effect on airborne diseases. McKeown clearly understood that clean water and pasteurized milk were important to improved health, but he noted that these benefits came late in the historical era of mortality declines. Supporting his claim about the timing of public health, consider that the role of germs was not understood until the mid-1800s, about the time that public health came into being, and pasteurization of milk did not start until around 1870 and its widespread commercial use did not come until well into the twentieth century. Having eliminated everything else, in his reasoning McKeown assumed that the great benefactor that transformed the developed countries from high mortality to low mortality must have been improved nutrition.

This argument for the primacy of nutrition, however, provided no direct evidence that nutrition improves health. Robert Fogel (2004) provided that needed evidence. He established that after the mid-eighteenth century, calorie intake of Europeans increased tremendously. At about the same time, their average height also increased substantially. The relationship of height to health is now well known; the Waaler Curve established that, for any given body mass, taller people (up to a point) have greater life expectancy (Fogel, 2004). We also now understand how better nutrition makes an individual better able to resist infectious disease. Fogel went on to study in great detail the heights and records of Civil War soldiers in the United States. His research led him to claim that nutrition played the major role in his book *The Escape from Hunger and Premature Death, 1700–2100*.

PUBLIC HEALTH REDUCED MORTALITY Other historical analysts take issue with the proposition that nutrition was the main cause of the mortality reductions. The crux of the issue is when the era of mortality reductions began. Public health advocates claim, contrary to McKeown, that the major declines did not start until around 1870, and if they began this late, then public health, which began about 1850, would have come in time to contribute. We know that the era from 1870 to about 1940 completed the "epidemiological transition." This phrase describes the remarkable transition in developed countries from when infectious disease was the major cause of death to a time when it became of only minor importance to population health. It is instructive to examine what public health accomplished during this period.

By 1870, cities had grown rapidly without the planning and development we now consider essential to a healthy environment. During this era, urban centers eventually and painfully slowly overcame their status of having worse mortality rates and general health than the countryside, the "urban deficit." Streets contained animal excrement, sewer systems were designed mainly for storm water, and water supplies were often delivered in lead pipes. The transition from water tainted with infectious organisms to clean water supplies was the most dramatic change in the health environments of city dwellers (see Box 5.2 for this story).

To summarize, the period from 1750 to the present contains three strands of health-related phenomena: (1) growth in life expectancy; (2) improved nutrition; and (3) improved public health. The difficulty is how to sort out which relationships proved most important. We see the importance of nutrition to body mass and height, the keys to health in the Waaler Curve (Fogel, 2004). This is compelling evidence. However, those who believe that public health was of key importance can point to the clean-up of cities, also compelling evidence. They point out, for example, that we know the modern Chinese people are not as tall as Americans, yet their life expectancies are much higher than would have been expected. Does not this point to the modern adoption of public health measures, which now can take place very rapidly? We will see shortly that these historical puzzles are not merely of "academic" interest but are critical for the growth in well-being of the lesser developed world.

BOX 5.2

The Importance of Clean Water

If transported by time machine back to the mid-nineteenth century, you would find it difficult to survive. This is because your modern body mass and height could barely be sustained by the small average quantities of available calories. But if you did survive and went to live in a city, you would find that in your weakened condition you would be very susceptible to infectious disease organisms permeating your environment, and especially in the water.

Even in 1900, waterborne infectious disease accounted for one-quarter of the deaths from infectious disease. Public health campaigns, which were painfully slow in gaining acceptance, cleaned up the water. They introduced the filtering of city water through sand. They fought to have sewage discharged at a safe distance from water intakes. In prior cases, cities had discharged waste directly into the same lakes or streams that provided drinking water. Water closets were introduced in about 1870, and these discharged human waste into a city sewer system that often could not handle it and overflowed even into the streets. Public health also introduced chlorination of the water supplies. If the earlier contaminating practices seem obvious and foolish to us, we need to remember that germ theory had only recently arrived, and pasteurization was discovered only in the late 1800s.

Cutler and Miller (2005) estimate that filtration of city water brought reductions in total mortality of 16 percent, and reduced infant mortality by 43 percent, in the 12 American cities studied. Applying cost-benefit principles, the researchers found that the ratio of benefits to costs in the filtration projects was about 23 to 1. This is history to us, but it is present-day reality to less developed countries, where over 1 billion people lack access to clean water. The United Nations has declared the 2005–2015 period the International Decade for Action on Water. More on the U.N. program can be found on the Web by searching for "millennium development goals."

What Lessons Are Learned from the Medical Historian?

We cannot conclude that medical research is unimportant in history or in the present day. Medical research contributes not only through improvements to medical practice, but also through its influence on health-enhancing practices. Typhoid provides a good example. Mortality from typhoid declined substantially well before the arrival by 1950 of chloramphenicol. Medical research, however, contributed to our understanding of the cause and transmission of typhoid and generated public health measures such as filtering public water supplies, chlorination of water supplies, and establishment of drinking-water standards. All of these factors occurred historically in time to have a major effect on mortality. Selma Mushkin (1979) estimates that medical research accounted for almost one-third of the cost savings to society from reduction in sickness or death rates in the United States from 1900 to 1975. The period 1900 to 1930 accounts for half the value of medical research effects, even though it came largely before the specific effective medical practice interventions.

Investments in medical research play a major role in our health and well-being. Murphy and Topel (2005) used people's willingness to pay for advances in medical knowledge as

a measure of its value. They estimated that the contribution of medical research to mortality reductions from 1970 to 2000 added $3.2 trillion to national wealth. Since 1970, reductions in heart disease mortality alone have been worth about $1.5 trillion. The medical research share of these gains, even if this amounted to only 10 percent of the total, easily compensates for the $36 billion we invest in it annually.

Second, perhaps the best result of this overview is a healthy skepticism toward the effectiveness of any given medical practice, and more importantly, to its significance and benefit to the population. It is in this spirit that the U.S. government has increasingly come to fund outcome studies. Outcome studies seek to address the effectiveness and appropriateness of specific medical practices on patient outcomes. The studies attempt to reduce the prevalent uncertainties in medical practice, and they offer important inquiries into the wisdom of using the marginal billion dollars on medical care delivery, particularly in terms of costs and benefits to the population as a whole. In the same spirit, "evidence-based medicine" aims to close the gap between outcomes research and physician practice.

Finally, and most importantly, these historical puzzles have relevance to the progress and public investment practices of lesser developed countries, who have scarce resources with which to invest in either industrial growth or to invest directly into health measures and public health improvements. If improved nutrition is the key to population health, then perhaps industrial growth will bring the best overall gains. If public health is the most direct and productive way to achieve health, it is also a route to develop a more productive workforce. These questions easily merit diligent research. We note also that there is an immense quasi-natural experiment to observe. Fogel (2000) and others (Pinkovski and Sala-i-Martin, 2009) report that world poverty rates are dropping sharply; the number of people in extreme poverty, those living on only $1 a day, has dropped sharply in the last two decades. We will want to see population health status in the developing countries improving.

The Production Function of Health in the Modern Day

The investigation of the modern health production function requires econometric techniques. An understanding of the strengths and limitations of these contributions requires attention to the underlying conceptual issues.

Preliminary Issues

Two conceptual issues bear on our interpretation of the results. These two issues can be posed as questions faced by every researcher: (1) how to measure health, the dependent variable in these studies, and (2) how to eliminate biases in the estimates.

HOW TO MEASURE HEALTH Consider the measurement of health. We desire a measure of population health status that captures the aspects of health status that are meaningful and that we can measure with adequate precision. It is difficult to attain both of these goals. We are most confident in the accuracy of mortality rate data, but mortality rates do not adequately capture several meaningful aspects of health status, such as reduction in pain and suffering and other improvements in the quality of life. The approach of past research in this field is understandable. Researchers have emphasized mortality data because of their accuracy, as well as because of their importance in the public mind. However, analysts have used other indicators of health status, such as morbidity (illness) rates and disability days. In discussing the empirical literature, we consider a variety of health measures.

ELIMINATING BIASES—REDUCED FORM VERSUS STRUCTURAL EQUATIONS
Consider that the statistical estimates often rely on *reduced form* equations, which are practical representations of the underlying true model of the phenomenon. The true model is based on what are called *structural* equations. Estimation based on the reduced form equations can lead to misinterpretations.[3]

The Contribution of Health Care to Population Health: The Modern Era

Health economists inevitably use different study designs and data sources to estimate the marginal product of health care. The resulting numbers need to be converted to a common basis, and for this we use the elasticity of health with respect to expenditure on health care inputs:

$$E_{Health:Expenditure} = \frac{\%\text{ change in health}}{\%\text{ change in health care expenditures}}$$

Table 5.2 reports the elasticities from several studies of the production of health. Each study applies econometric methods to analyze survey data; these range from statewide data to data on county groups (Hadley, 1982, 1988) to data on individuals (Sickles and Yazbeck, 1998). Though not shown in the table, recent investigations (Cremieux, Oulette, and Pilon, 1999) confirm a significant contribution of health spending to reduce infant mortality (in Canada). Similar reports also come from India (Farahani, Subrmanian, and Canning, 2010).

The several studies offer insights because of their differences. Some suggest that lifestyle and environment expenditures could provide more benefits per dollar of cost than health care; and efforts to improve schooling or reduce cigarette smoking offer appealing trade-offs with health care spending. The studies we have described help establish that health care spending makes a statistically significant contribution to health and argues that health care passes benefit-cost criteria at the margin. The latest study follows its subjects through time.

A study with both quality and novelty of design adds support to our finding that medical expenditures improve health (Almond et al., 2010). Low-birth-weight babies have a worse chance of surviving, and with "very low birth weight" defined as below 1,500 grams. It will

Table 5.2 Measuring the Contribution of Health Care to Population Health

Study Cited	*Date of Study*	*Health Care Elasticity**	*Significant?*
Hadley	1982	0.12 to 0.17	Yes
Hadley	1988	0.20 to 1.00	Yes
Sickles and Yazbeck	1998	0.03 to 0.05	Yes

Note: *This is the elasticity of health with respect to health care expenditure.

seem odd, but babies just below that cutoff had one percent *lower* mortality rates than babies just above it. The reason was that "very low birth weight" qualifies the baby for special medical treatment (with extra costs). The authors put these facts together finding that the marginal product of the extra care significantly contributed to health. They further concluded that the cost of saving a statistical life of a newborn with birth weight near 1,500 grams is of the order of $550,000 in 2006 dollars (Almond et al., 2010). Based on value of life estimates (see Chapter 4), these medical efforts are clearly worth it.

Is Health Care Worth It?

But are we on the "flat of the curve," getting statistically significant but very small marginal product from the marginal investment in health care? The answer is clearly no. Murphy and Topel (2005), as we have seen, estimate that the American gains in life expectancy from 1970 to 2000, based on willingness-to-pay principles, are worth over $3 trillion. We invest a small fraction of that in health technology research. Much of the gain is probably due to health care. Similarly Cutler (2004) ties 40 percent of the gains in life expectancy between 1950 and 2000 to medical improvements in just two categories: low-weight infant care and cardiovascular disease treatment. Again, based on willingness to pay, the life gains are worth the extra costs of medical care overall—not even counting the benefits from other forms of health care.

Look again at the marginal benefit of health care (technology held constant), which is small but not zero. We earlier suggested that the health production elasticity of health care on the margin is about 0.10. To illustrate the meaning of this elasticity, suppose that 0.10 were the true production elasticity of health care, and let Congress reallocate $250 billion from other programs to health care expenditure. In 2011, the United States spent $2.5 trillion on health care. The $250 billion transfer would increase health care expenditure by 10 percent. We would extrapolate the improvement in health to be $10 \times 0.10 = 1$ percent. If we define "health" by average life expectancy, a 1-percent gain would mean an increased average life expectancy of $78 \times 0.01 = 0.78$ year, provided this increased expenditure continues indefinitely. Spread over the population (for a life span), however, the 0.78 year of life could lead to incremental benefits that exceed the incremental costs.

BOX 5.3

Sulfa: A Drug That Really Made a Difference

At a time when cost-control planners seek to sort out medicine that is expensive but not very effective, sulfa is a good example to remind us that there are medicines and practices that work extraordinarily well. It was the miracle drug of the 1930s, made available in 1937 several years before penicillin appeared. Sulfa was a major player in the epidemiological transition from a time when infectious diseases were the fearsome killer to modern times when non-communicable diseases predominate: The following example provides an excellent way to measure sulfa's effect.

Sulfa was not patented, appeared suddenly, and was distributed rapidly. These facts made the case ideal for study by health economists Jayachandran, Lleras-Muney, and Smith (2010). They tested population health effects of sulfa by comparing the

time patterns of those mortality rates of diseases treatable by sulfa with those that weren't. In most of these cases, the results were dramatic. Sulfa caused sharp reductions in mortality for treatable diseases such as MMR (measles, mumps, and rubella), pneumonia, scarlet fever, and maternal mortality. The drug lowered maternal mortality from 60 percent to 36 percent and caused similar or higher drops in the other sulfa-treatable diseases. The most telling statistic is that, by itself, sulfa raised U.S. life expectancy by 0.4 to 0.7 years.

On the Effect of Social Health Insurance

The early "standard" study was the RAND Health Insurance Experiment (RHIE), one of the largest randomly controlled economic experiments ever conducted. It was designed to test the effect of alternative health insurance policies on the demand for health care and on the health status of a large and closely observed group of people from all walks of life.

RAND researchers discovered that the greater the portion of the health care bill that individuals are required to pay, the less health care they choose to purchase. Fortunately, RHIE analysts kept detailed records on each person, including a dozen or more measurements under each category of physical health, mental health, social health, and general health index. They also examined their subjects' dental health, persistence of symptoms, health habits, and disability days. The results are easy to summarize. For dozens of items, virtually no differences were found between the groups studied; health care and health insurance did not seem to matter.

A simple example from the RHIE illustrates the point. Table 5.3 provides detail on work-loss days per employed person per year—a measure of health status and morbidity that some economic researchers like to use because it ties directly to both health and productivity. This table separates the RAND subjects into four groups, which differ by type of health insurance policy. Some subjects pay nothing out-of-pocket for their health care/health insurance package; some pay 25 percent to 50 percent of their bill themselves; others pay all of their health care bills up to a certain amount, called a deductible. The subject's out-of-pocket cost ranges from zero (free) to about 95 percent of the bill. Newhouse et al. (1993) summarize: "Our results show that the 40 percent increase in services on the free-care plan had little or no effect on health status for the average adult."

The effects on children showed a somewhat similar pattern. Valdez et al. (1985) examined data for 1,844 children in the RAND study—children who differed primarily by the type of insurance plan their families obtained. Children under the cost-sharing plans consumed up to one-third less care. However, the reduction in care was not significantly related to health status measures.

It may seem from the RAND results that public provision of health insurance to both adults and children might not be justifiable on the basis of benefits to health. However, as Jonathan Gruber (2008) points out, this conclusion does not follow. No one in the RAND Experiment was "uninsured," completely without insurance, as are close to 30 million Americans as of this writing. The least insured individuals studied by RAND had full coverage for health expenditures above a deductible, which was $1,000. Studies of the truly uninsured began to appear showing significant health gains from the provision of public insurance (Currie and Gruber, 1996; Doyle, 2005; Hanratty, 1996). These studies report

Table 5.3 Work-Loss Days per Employed Person per Year, by Plan

Plan	*Mean*	*Standard Error of Mean*	*95% Confidence Interval*	*Number of Persons*
Free	5.47	0.42	4.65–6.29	1,136
Intermediate (25%, 50%)	4.82	0.37	4.09–5.55	983
Individual Deductible	4.54	0.36	3.83–5.25	787
Family Deductible (95%)	4.82	0.53	3.78–5.86	600

Source: Reprinted by permission of the publisher from *Free for All? Lessons from the RAND Health Insurance Experiment* by Joseph P. Newhouse et al., Cambridge, MA: Harvard University Press, 1993. Copyright © 1993 by the RAND Corporation.

reductions in infant and neonate deaths of around 5 to 10 percent. The Institute of Medicine estimates suggest that even larger gains are possible; they claim that the uninsured face a 25 percent greater mortality risk.

The pursuit of the social health insurance effects on health takes on a new energy because of new experiences with the Massachusetts Health Reform 2006–2011, The Affordable Care Act (2010), and the unusual Oregon expansion of Medicaid. Of these the ACA is too new to analyze health effects, and we describe the Massachusetts results shortly. But consider how the Oregon case is "unusual." The Oregon administrators had arranged to give access to Medicaid by lottery, making the expansion random and thus scientifically comparable to the RAND experiment.

So far we have the studies of the first two years. Researchers have concluded through this randomized controlled study that Medicaid coverage generated no significant improvements in measured physical health outcomes in the first two years, but it did increase the use of health care services, raise rates of diabetes detection and management, lower rates of depression and reduce financial strain (Baicker et al., 2015; Finkelstein et al., 2012).

The Massachusetts Health Care Reform lacks the randomized structure of the RHIE or the Oregon study but it has two features giving it substantial interest to health economists: first it ran for more years, and second it has the same basic design as the ACA (Obamacare). Its reported results are somewhat more optimistic. Van der Wees and colleagues (2013) find that after health care reform, Massachusetts residents reported better general, physical and mental health, than residents from neighboring states.

Prenatal Care

The importance of examining population subgroups comes from neonate mortality studies (for example, see Corman and Grossman, 1985; and Corman, Joyce, and Grossman, 1987). A neonate is an infant one month old or younger. Thus, the neonate mortality rate refers to deaths to neonates per 1,000 live births. How can we reduce these deaths? The early studies observed counties in the United States and identified several factors that seemed to cause higher neonate mortality rates. Table 5.4 displays the production of neonate mortality study by Corman, Joyce, and Grossman (1987).

Table 5.4 Contribution of Selected Factors to Reductions in Neonate Mortality Rates, 1964–1977

Factor	*Whites Total Effect*	*Blacks Total Effect*
Organized family planning	0.084	0.526
WIC	0.425	1.330
BCHS	0.002	0.030
Neonatal intensive care	0.140	0.534
Abortion	0.824	2.109
Prenatal care	0.434	1.949
Total explained reduction	1.9	6.5
Total reduction	7.5	11.5
Percentage explained	25.3	56.5

Note: Figures record estimates of the reduction in deaths per 1,000 live births predicted to have been caused by various factors.

Source: Reprinted from *Journal of Human Resources*, Hope Corman, Theodore J. Joyce, and Michael Grossman, "Birth Outcome Production Function in the United States," *Journal of Human Resources* 22 (1987): 339–360, with permission from The University of Wisconsin Press.

The table presents their estimates of the contribution of each measured factor to the reported mortality rate decline. The WIC program is the shortened name for the governmental program designed to provide improved nutrition for women, infants, and children; it is a means-tested program, meaning that it is directed to the poor. The BCHS variable is the authors' measure combining various Bureau of Community Health Services projects, including maternal and infant care, as well as community health centers.

The data recorded in Table 5.4 indicate that of the total reduction in neonate mortality for whites during the period, 1.9 deaths per 1,000 live births, or 25.3 percent, can be explained by the observed factors, whereas for blacks a greater amount, 6.5 or 56.5 percent, can be explained. Blacks benefit more from health care on the margin, a finding that has been found in other studies. Interpret the table as follows: During the period studied, the WIC program resulted in a reduction of white neonate mortality rates of 0.425 deaths per 1,000 live births, while for blacks WIC reduced neonate mortality by 1.330 deaths per 1,000 live births.[4] Note too, that abortion tends to lower mortality; this result may reflect that many abortions were of fetuses that would not have survived infancy. Also note that prenatal care is effective in these data, especially for blacks; prenatal care shows up well in most studies.

A novel experiment (Evans and Lien, 2005) reinforces the value of prenatal care. When public transportation workers in Allegheny County, Pennsylvania, went on strike in 1992, prenatal visits among the poor were sharply curtailed. The results were significant reductions in the quality of birth outcomes, especially for those affected during the early stages of pregnancy.

Aizer, Currie, and Moretti (2007) observed the value of prenatal care indirectly. They measured the effects of a law transferring some California Medicaid patients to managed care, which in this situation offered distinctly lower quality prenatal care. Significantly greater numbers of low-weight births were observed. Another study, by Conway and Kutinova (2006), finds prenatal care to be effective in reducing the probability of low-weight births.

The World's Pharmacies

Note also the contribution of pharmaceutical availability in determining population health. Recent research (Shaw, Horrace, and Vogel, 2005; Miller and Frech, 2004) finds that countries in the Organization for Economic Cooperation and Development (OECD) with higher drug consumption have greater life expectancies. The magnitude of the effect is on a par with the effects of reduced cigarette consumption and increases of fruit and vegetables in the diet.

Do Other Measures of Health Care Affect Health?

On the Importance of Lifestyle and Environment

Didn't we always know that much of our health depends on the wisdom of our own choices? The role of lifestyle was best illustrated by Victor Fuchs in his book *Who Shall Live*? (1995). He compared average death rates in Nevada and Utah for 1959 to 1961 and 1966 to 1968. These two states are contiguous, and they share "about the same levels of income and medical care and are alike in many other respects" (p. 52). Nevertheless, average death rates in Nevada were greater than those in Utah. Table 5.5 shows the

Table 5.5 Excess of Death Rates in Nevada Compared with Utah, Average for 1959–1961 and 1966–1968

Age Group	*Males (%)*	*Females (%)*
Less than 1	42	35
1–19	16	26
20–39	44	42
40–49	54	69
50–59	38	28
60–69	26	17
70–79	20	6

Source: Reprinted from Victor R. Fuchs, *Who Shall Live? Health, Economics, and Social Choice*, Expanded Edition, Singapore: World Scientific Publishing Company, 1995, p. 52, with permission from the author and World Scientific Publishing.

results of Fuchs's work. Fuchs argued that the explanation for these substantial differences surely lies in lifestyle:

> Utah was, and remains, inhabited primarily by Mormons, whose influence is strong throughout the state. Devout Mormons do not use tobacco or alcohol and in general lead stable, quiet lives. Nevada, on the other hand, is a state with high rates of cigarette and alcohol consumption and very high indexes of marital and geographical instability.
>
> (p. 53)

In 2009, Utah, with its low age-adjusted death rates, was still a national leader in health (this death rate equaled 507.8), while Hawaii (717.9) and Nevada (727.3) were much higher, but significantly lower than the national average (793.7). Before concluding that a simple life and plenty of sun are the tickets to good health in and of themselves, consider that many of the top 10 healthful states, while they may be sunny, are known to be chilly: Minnesota (718.6), New Hampshire (761.6), Idaho (774.5), and Colorado (620.3). (Data source: CDC, Preliminary death rates, 2009.)

Cigarettes, Exercise, and a Good Night's Sleep

Many have chosen to quit smoking (or to avoid becoming addicted to cigarette smoking in the first place). Americans know that heart disease and cancer are the two leading killers, but most do not realize how substantial a part smoking plays. Using the category "malignant neoplasms of the respiratory system" (the category for lung cancer), we find that the 2002 death rate (51.5) is twice as high as that for any of the following: breast cancer (13.4), prostate cancer (9.2), pneumonia and influenza (17.5), diabetes mellitus (22.3), HIV (3.1), or motor vehicle-related injuries (11.8).[5] We already have seen the negative health production elasticity of cigarettes, which makes it clear that cigarette smoking affects the average health of the community and is statistically significant at that level.

However, economics searches for underlying causes, and human behavior can have many interwoven causes. For example, smoking and other lifestyle behaviors may themselves be determined by unobserved variables that affect health status. This common problem in economic empirical work has been addressed in recent research (Balia and Jones, 2008; Contoyannis and Jones, 2004). They address the problem by estimating both the determinants of lifestyle behaviors as well as the determinants of health status, giving a clearer picture of the importance of lifestyle. The authors showed that a good night's sleep, avoiding smoking, and regular exercise each contribute importantly to self-reported health.

While smoking certainly causes ill health, it is pleasurable as well, and one's degree of health can affect the decision to quit. For example, a healthy individual may be more likely to quit as a preventive measure; on the other hand, a critically ill individual may quit as a curative measure (Jones, 1996). Folland (2006) shows that greater life satisfaction means being less willing to risk death by smoking.

Granted that lifestyle is a major player in health comparisons between individuals, it is natural to ask whether it plays the same role when comparing countries. As we have seen earlier in the course, life expectancy in America is lower than in many developed countries. Commanor, Frech, and Miller (2006) investigated this question. They began by assessing U.S. efficiency in the production of health, finding it to be somewhat less efficient than other

developed countries. What is most relevant to our present discussion is their finding that much of the U.S. deficit stems from the higher rates of obesity in the United States.

The Family as Producer of Health

Women have long been warned to avoid cigarettes and alcohol while pregnant. Are such lifestyle factors important enough to be included as factors in the production of newborn health? The answer is yes. In the production of newborn birth weight (an important guide to infant health outcomes), maternal cigarette smoking has a significant negative effect (Rosenzweig and Schultz, 1983; Rosenzweig and Wolpin, 1995). Data on maternal smoking now show that taxing cigarettes leads to improved birth outcomes via its effect on smoking behaviors of expectant mothers (Evans and Ringel, 1999).

Maternal behavior also can have strong and tragic consequences in the case of drug use. Joyce, Racine, and Mocan (1992) found that the alarming increase in low-birth-weight births in New York City, particularly among blacks, was due in large part to an epidemic of illicit substance abuse by pregnant women. The explosion of cocaine use had horrible consequences for these babies.

Looking at this at a more abstract level, a study from Sweden (Bolin, Jacobsen, and Lindgren, 2002) develops the theory of how parents make health investments in themselves and their children. If parents individually make these health investment decisions strategically—that is, in response to the expected decisions of the others—the decisions, together, will not be optimal for the family. Even more significant health investment problems will occur, they warn, when parents split up in divorce, because the non-caregiver may lose some incentives to invest in the child's health. The parent's incentive to invest in the children's health is clearly a critical factor in child health.

Social Capital and Health

Recent research has made it clear that family, friends, and community are associated with the health of the individual and the community. The networks of social contacts of an individual or the complex overlapping networks in a community have come to be called *social capital*. The effects, first described by political scientists, sociologists, medical researchers, and epidemiologists, suggest that social capital beneficially affects measures of health (see Islam et al., 2006, for a review).

Social capital may improve an individual's health in several ways: (1) it may relieve stress to have the support of more social contacts; (2) more contacts can provide additional information on healthful behaviors and health purchases; and (3) satisfying social relationships may provide reasons to re-evaluate risky health behaviors. This issue presents complex research obstacles; for example, not all social contacts are beneficial.

The bigger issue, however, is how to determine whether social capital in these studies causes better health or alternatively whether it is a result of some other factors. This is an important avenue by which economics and its econometric tools provide benefits to the ongoing research of other disciplines.

Health economists have taken interest in this area as a potential subject area in which to make a joint contribution with other disciplines. This is occurring both in theory and in empirical work.[6] Findings generally support the hypothesis that social capital improvements lead to health improvements.

Environmental Pollution

Pollution causes ill health and death in individuals, with the elderly and people with respiratory diseases more susceptible. The degree to which reductions in pollution will improve the health of populations is somewhat less clear. Pollution effects on health are sizable and statistically significant in both industrialized and lesser-developed countries (Cropper et al., 1997). Based on levels of total suspended particulates (TSP) in New Delhi between 1991 and 1994, the average pollution level was five times the limit recommended by the World Health Organization (WHO). Variations in deaths in New Delhi responded statistically to the variations in pollution; if these estimates prove true, then a reduction of pollution levels of about one-third would reduce deaths by more than 2 percent.

A similar study by Schwartz and Dockery (1992) in Philadelphia suggests that reducing the pollution level there by the same 100 micrograms per cubic meter would reduce deaths by more than 6 percent in the general population and nearly 10 percent for the elderly. This is because with our generally better health status in the United States, more people live long enough to become part of the population most sensitive to respiratory problems from pollution.

Income and Health

While we know that good health during the years when an individual is forming a career can be a big boost to that person's income later in life (James Smith, 1998), we also know that being rich does not necessarily cause one to choose to live and eat wisely. Even programs designed to raise the income of poor families, such as (the late twentieth century) Aid to Families with Dependent Children in the United States, did not always correlate with good health habits among the recipients (Currie and Cole, 1993; Currie and Gruber, 1996).

Though earlier work had suggested that being richer in America was generally better for one's health, research by Deaton and Paxson (2001) brought that conclusion into question. Examining in detail both U.S. and British data over time, they find the relation of income and health to be complex and contradictory. There was a substantial decline in mortality after 1950, but rather than growing incomes as the cause, they conclude "a more plausible account of the data is that, over time, declines in mortality are driven by technological advances, or the emergence of new infectious diseases, such as AIDS" (p. 29).

Part of our problem thus far in researching the contribution of income to health in the industrialized world is that incomes do not vary greatly enough to detect the larger patterns. Pritchett and Summers (1996) leave little doubt that extremely low incomes have a strong effect on people's health. Though they treated the econometric challenges with great respect in establishing their conclusion, the most persuasive arguments may be those provided by simple graphs fitting various health statistics to per capita income data. These curves fit well and reveal that "modern" standards of good health are enjoyed solely by the industrialized countries with mortality experience turning sharply worse with lower income levels, conditions common in the underdeveloped world.

The Role of Schooling

What is the role of education? Since education includes both formal and informal training (such as experience or on-the-job training), some portion of education is impossible to

measure accurately. Most often health economists focus on schooling as measured by years of schooling, or academic accomplishments such as diplomas or degrees. Health status correlates significantly with schooling as we have seen. If the marginal product of health care is truly small ("we are on the flat of the curve"), then perhaps we should reduce public health expenditures on health care at the margin and transfer the expenditures to education. However, the wisdom of such a policy depends on which of the two theories is correct.

Two Theories about the Role of Schooling

The ideas and work of two health economists serve to develop and contrast the two theories. First, Michael Grossman's (1972a, 1972b) theory of demand entails a central role for education. Grossman contends that better-educated persons tend to be economically more efficient producers of health status.

In contrast, Victor Fuchs (1982) has suggested that people who seek out additional education tend to be those with lower discount rates. A decision maker with a high discount rate will tend to prefer projects with immediate payoffs versus long-term projects. People with a lower discount rate tend to be those who value the long-term gains more. Now consider individuals facing a possible investment in education. Because education requires current costs to gain distant payoffs, individuals with relatively low discount rates will be more likely to invest in education and in health as well.

Empirical Studies on the Role of Schooling in Health

Recent evidence supports the view that education makes one a more efficient producer of health (Lleras-Muney, 2005). Knowing that compulsory education laws came into being in various places at various times in the twentieth century, she reasoned that the related birth cohorts from that era would have experienced different levels of education but would have been similar in many other respects. This formed a natural experiment in which she could analyze the survival patterns of these people to detect a pure influence of education on health. Furthermore, the education laws could not have been directly manipulated by the study subjects, so they were good "instruments" for education. By this approach, she was able to conclude that education has a clear, causal, and positive effect on health. By 1960, the early century education experience appeared to have increased life-years by 1.7 years, a substantial increase and one not due to time preferences of the subjects. See also Webbink, Marti, and Vischer, 2010; and Amin, Behrman, and Spector, 2013.

Lleras-Muney's study inspired new research on the effects of new laws extending the length of compulsory education in England and Ireland (Oreopoulus, 2006; Auld and Sidhu, 2005). These supported the earlier findings; an additional year of schooling caused an improvement in the affected student's health. To emphasize, the improved health was experienced by "likely dropouts," forced by the law to attend one more year of high school. Interestingly, Lindeboom and colleagues (2006) inquired through research as to whether the children born later on to these students *also* benefited from improved health, but the findings indicated that they did not.

In summary, research has supported the theory that education makes people more efficient producers of their own health. Cutler and Lleras-Muney (2006) add further support in their recent work by showing that education is associated statistically with better reasoned choices of health-related behaviors. One finds as well that education plays a stronger role in health for cases where new medical knowledge is more important.

Conclusions

In this chapter, we investigated many topics related both directly and indirectly to the production of health. The health production function exhibits the law of diminishing marginal returns. While the total contribution of health care is substantial, the marginal product is often small. Historically, we found that much of the decline in mortality rates occurred prior to the introduction of specific, effective medical interventions. Thus, historically the contributions of health care, at least as provided by the health practitioner, were probably small until well into the twentieth century. The small, modern-day marginal product of health care is statistically significant. Health care benefits people differentially and is generally more productive on the margin for women and blacks. Similarly, certain categories of health care have greater marginal effects on the population than others; prenatal care programs are examples of the more productive categories. Education has a strong association with health status. Whether this means that it causally improves health has long been an issue of contention. Recent research supports the view that education improves health.

Summary

1 The production function for health exhibits diminishing marginal returns. In developed countries, the total product of health care is probably substantial at the same time that the marginal product is relatively small.
2 The historical declines in mortality rates in representative industrial countries were substantially responsible for the large growth of populations.
3 The historical declines in population mortality rates were not due to medical interventions because effective medical interventions became available to populations largely after the mortality had declined. Instead, public health, improved environment, and improved nutrition probably played substantial roles.
4 The marginal product of health care in the United States is small. Recent studies find elasticities in a narrow range around 0.10. Nevertheless, the total contribution of investment in health care technology over the past several decades is probably in the trillions of dollars; much of the improvement in life expectancy in this period can be attributed to health care improvements.
5 The RAND Health Insurance Experiment found that increased use of health care has little effect on the illness rates of the study population. However, studies of the totally uninsured now reveal gains in health due to publicly provided health insurance.
6 Lifestyle and environment are major and statistically significant determinants of population health status.
7 Health care contributes more substantially to health for subgroups of the population, including infants and also certain ethnic minority groups.
8 Social capital, produced by groups, is increasingly viewed as a substantive determinant of individuals' health.
9 Education, as measured by years of schooling, is positively related to population health. Recent research supports the view that the relation is causal, that increased education improves health.

Discussion Questions

1 Assume that health production is subject to diminishing returns and that each unit of health care employed entails a constant rate of iatrogenic (medically caused) disease. Would the production of health function eventually bend downward? Explain.
2 What evidence is there to suggest that the United States is on the "flat of the curve" in health production? Is a typical developing country likely to be on the flat of its health production function? Discuss the differences.
3 Which of the following are important in explaining the modern rise in population in England and Wales: birthrates, death rates, and net migration rates? Describe the evidence.
4 "Medical interventions were not important in the historical declines in mortality rates, but that does not imply that medical research was unimportant." Explain this viewpoint.
5 What role did public health play in the historical decline in mortality rates?
6 Suppose you were hired as an adviser to a developing country and you were versed in the theory of production, the historical role of medicine, and the modern-day health production function studies. Their government seeks advice on the wisdom of a relative emphasis on health and health investment versus other forms of economic investment. What would be your advice?
7 Someone says the following: "Lifestyle may be the most important determinant of health status, but changing lifestyles may not be the least costly way to improve population health status." Explain the circumstances under which this opinion could be true. Is it likely to be true in reality? What does the evidence on lifestyle suggest about government policies to improve the public's overall health?
8 Summarize the two theories on how schooling is correlated with health status. Which of the two theories does the evidence support?
9 Research shows that the returns for prenatal health care are high, whereas it may cost hundreds of thousands of dollars to keep an acutely ill, elderly person alive. What does this suggest about the appropriate allocation of resources among members of society?
10 What is the total contribution of health care to health as estimated by researchers? What parts of health care are most effective in this regard?

Exercises

1 Graph the production of health function $HS = 10HC^{0.5}E^{0.3}\ LS^{0.4}HB^{0.2}$ in a graph with axes HS and HC, assuming $E = 10$, $LS = 5$, and $HB = 7$. Graph the marginal product of health inputs. Is it increasing or decreasing? Show how the curve changes when E is increased to 15.
2 Which factors in Table 5.4 were important in explaining improvements in black neonate mortality rates? White neonate mortality rates? Speculate on why some of these factors may have been more important for blacks.
3 What are the differences between mortality and morbidity? Would you expect the two variables to be related to each other? If so, how?
4 We know that correlations never explain; it is our theories that provide explanations. Re-examine Table 5.4 and draft theories to explain why WIC has a larger contribution

than Organized Family Planning. Why do the contributions differ between blacks and whites on WIC, abortion, and prenatal care?

5 Pritchett and Summers argue that income per capita is strongly and positively related to health status when viewed across the world. From data in Table 2.2, in Chapter 2 of this text, plot a graph of GDP per capita against life expectancy for the countries shown. Does your plot confirm the Pritchett and Summers finding?

Notes

1 This categorization of groups of inputs is not the only reasonable one, but it illustrates the main issues, and it has an excellent history. It was developed by the Canadian government for its pioneering work begun over four decades ago (Lalonde, 1974).

2 The medicalization argument was reinforced by Thomas McKeown's (1976) research showing that medical care warranted little credit for the historical declines in mortality rates in England and Wales. The works by McKeown and by Cochrane remain the reference points for the continuing debate within medicine and the social sciences over the role of medicine (Alvarez-Dardet and Ruiz, 1993; Farmer and Nardell, 1998; Frenk, 1998).

3 For further discussion of this issue and method, see Rosenzweig and Schultz (1983); Grossman and Joyce (1990); and Atkinson and Crocker (1992).

4 Currie and Gruber (1996) showed general Medicaid eligibility to improve birth outcomes in a study across states in the United States. (This is distinguished from the particular Medicaid spending directed to prenatal care.) These authors, however, raised issues about the cost-effectiveness of improvements in Medicaid eligibility.

5 These death rates are age-adjusted deaths per 100,000 resident population, National Vital Statistics System, 2009.

6 Folland (2006, 2008); Folland, Kaarboe, and Islam (2011); Folland and Rocco (2014).

Chapter 6

The Production, Cost, and Technology of Health Care

In this chapter

- Production and the Possibilities for Substitution
- Costs in Theory and Practice
- Technical and Allocative Inefficiency
- Technological Changes and Costs
- Diffusion of New Health Care Technologies
- Conclusions

Recognizing that health is the ultimate output in the health sector, we understandably direct considerable policy interest toward the production and cost of the intermediate output, health care. Despite exceptions, both politicians and consumers seeking lower health costs do not seek cuts in "necessary" health care—only that which is "purely wasteful." Yet, health care, even when needed, inevitably "costs too much."

This chapter addresses production, cost, technology, and efficiency of health care. We investigate first the question of what degree of flexibility the manager or the public policymaker can find in the production process in order to make improvements. Must we always call physicians for certain tasks, or can we substitute nurses and other less expensive inputs? Can managers safely vary the mix of types of nurses employed? These questions reduce the degree to which health care inputs substitute for one another.

Second, the study of cost functions can provide further clues to finding efficiency gains. Economists find that economies of scale and scope exist in many industries, and society would be better off if firms chose the size that minimizes average costs. Theory states that the perfectly competitive industry achieves this in the long run without outside interference. Health care firms are generally not perfectly competitive nor necessarily perfectly managed, and health economists inquire into the extent of any excess costs.

The next issue of the chapter is the technical and allocative efficiency of health care firms. These types of efficiency and the inefficiency they define articulate in economic terms the central research issue of the American consumer's complaint: "I am paying an arm and a leg for my family's health expenses. Am I really getting my money's worth?"

We then examine how greatly the picture changes as technology, the major mover in the health care industry, changes. We will see that even when new health technology improves our lives, it can also make life more expensive. Insurance plays an important role as a shifter of demand, though probably its influence through its effect on technological innovations is as important.

Finally, we study the economics of how and why new health care technology diffuses through the health system and forms a "logistic" pattern. The diffusion can be rapid, although it may be slowed by regulatory or institutional realities.

Production and the Possibilities for Substitution

Economists often note that there is more than one way "to skin a cat," that different techniques are available to produce the same product. A single technique is one recipe for production, meaning one specific combination of inputs. When multiple techniques are available, one can choose a relatively capital-intensive (labor-intensive) technique during times when capital (labor) is relatively cheap. It also means the ability to use cheaper forms of labor in substitution for more expensive forms.

While multiple techniques are common in many industries, health practitioners often recognize only one correct way of treating a given illness. The belief that only a single technique is possible or wise is what Victor Fuchs has called the "monotechnic view." If such a view correctly described production processes, cost-saving substitutions would be difficult if not impossible without reducing either output or quality. A more flexible production process permits cost-saving improvements that may be beneficial to consumers. Economists investigate this question using the concept of substitution.

Substitution

Flexibility means the ability to substitute one input, such as capital, for another input, such as labor, while maintaining the level and quality of output. This does not mean that the

two inputs are equivalent, but only that alternative combinations are possible. Figure 6.1 illustrates a case with no input substitution and a case in which an infinite number of techniques are available. In panel A, the isoquant shows the possible combinations of nurse hours and physician hours required to treat one patient case in a hospital; the isoquant is labeled $Q = 1$. Given this situation, only one sensible production technique combines the two inputs. Physicians and nurses must be combined in the ratio given by $0P/0N$, the ratio of inputs used at the corner point M. Notice that $0P/0N$ is also the slope of line segment $0M$.

What does this mean? In panel A, $0P$ physician hours are required to produce one case, and the addition of nursing hours beyond $0N$ will not add to output unless physician hours also are increased. This applies to a production problem where patient care requires certain professional tasks that only a physician is trained and competent to perform.

The fact that the isoquant is flat when moving to the right from M means that adding nurses beyond the required combination produces no more output (i.e., they would be wasted). Likewise, the fact that the isoquant is vertical when moving upward from M means that additional physician hours beyond the required ratio combination are simply wasted resources. How would an isoquant look if substitution were possible? Panel B illustrates this. Again, a unit isoquant is depicted, but a smooth downward-sloping convex curve illustrates that many combinations of inputs could be chosen without being wasteful. For example, one case can be treated with the $(0P, 0N)$ combination of inputs or equivalently with the $(0R, 0S)$ combination. Each point on the isoquant represents a different technique. The slope of any isoquant, such as the isoquant labeled $Q' = 1$, is called the marginal rate of technical substitution (see Chapter 2), and it represents the rate at which nurse and physician hours can be exchanged while still maintaining output.

We emphasize that even though we can substitute in panel B, nurses and physicians are not equivalent. It is not even true that a fixed number of nurses always can replace a physician. Moving along the curve from point Y to point Z, the rate of substitution changes; that is, the slope becomes flatter, indicating a diminishing marginal rate of technical substitution.

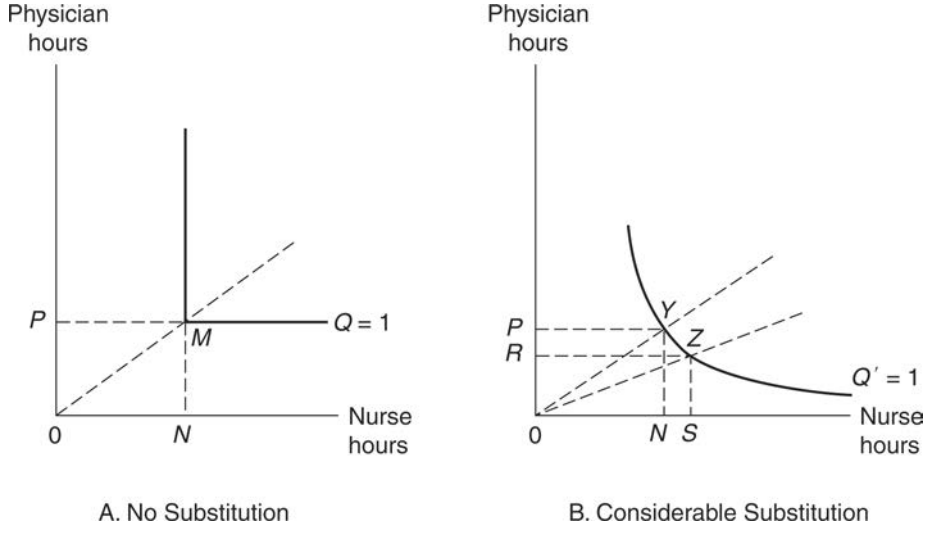


Figure 6.1 Degree of Substitution between Physicians and Nurses

To replace one hour of physician time at point *Y* requires some number of nursing hours; however, at point *Z*, where we are using fewer physician hours, we require a much greater number of nursing hours. That is, as physician time grows scarcer relative to nursing hours, it becomes more difficult to replace. This retains the idea that while substitution is possible, it may be difficult, expensive, or unsafe to have nurses do certain physician tasks, and vice versa. The curve may even become flat at some point, indicating that we have reached a minimum of required physician time.

What Degree of Substitution Is Possible?

The graphs in Figure 6.1 define terms, but they do not tell us which of the two cases is true of the health care world. From the 1970s on, economists addressed the question of whether physician extenders (as noted in Box 6.1) could substitute for physicians. Physician extenders refers to specially trained physician assistants or nurse practitioners who are utilized to perform certain tasks, including some that formerly were performed by the physician.

The estimates obtained are equivalent to the measurement of a few points along an isoquant, such as the ones just depicted. This work suggested that substitution could be substantial. Depending on the number of physician hours employed, one physician extender could substitute for 25 percent to more than 50 percent of a physician's services. Since physician extenders have a much lower training cost, this degree of substitutability could result in substantial savings. Reportedly, the observed physicians were not employing sufficient numbers of assistants to make the practice fully profitable to the physicians.

Elasticity of Substitution

The hospital has provided a related focus of production studies. One study presents evidence of substitution among various categories of hospital inputs. It also illustrates a second and more convenient way to measure the possibilities for substitution between inputs, the elasticity of substitution (E_s), which measures the responsiveness of a cost-minimizing firm to changes in relative input prices. It is defined as follows:

$$E_s = \frac{\text{Percentage change in factor input ratio}}{\text{Percentage change in factor price ratio}}$$

What does this elasticity mean? If a firm were a cost minimizer, then it would be responsive to changes in input prices, and it would tend to respond by shifting away from the now costlier input to the now relatively cheaper input. Suppose that a hospital is currently at combination *Y* in panel B of Figure 6.1 using 0*P* physician hours and 0*N* nursing hours to treat one case. Suppose also that physicians are paid $200,000 per year and nurses are paid $40,000 per year. Suppose, finally, that the hospital employs 100 physicians and 100 nurses. If there is a 10 percent increase in the relative wage rate of physicians (from $200,000 to $220,000), the cost-minimizing hospital moves to combination *Z*, which substitutes *NS* nursing hours for *PR* physician hours. The relative physician input ratio decreases from the one shown by the slope of the line segment 0*Y* to the one shown by the slope of 0*Z*.

BOX 6.1

Health Care Professionals: Expanding the Possibilities

Most research on health professionals tends to focus on physicians, dentists, and managers. Yet each researcher is aware that a large share of "getting better" depends on the allied health professions. These professionals, besides often providing a human touch to care, can significantly improve the health production process.

First, the variety of specializations among allied health labor provides the benefits of Adam Smith's "division of labor." The theory explains how a focus by each expert on his or her specialized task yields greater output for the whole than were one to insist that each person be a generalist. Imagine a hospital where every professional was a generalist. This health professional then keeps the electronic records, prepares patients, sees patients, draws blood, takes X-rays, and so on. With present-day complex technologies, this is not even possible, let alone efficient. Office managers, registered nurses, medical technicians, X-ray technicians, physicians, and medical records specialists together solve this economic problem as a team.

Second, the availability of other health professionals enhances the possibilities for substitution in production. Nurse practitioners and physicians' assistants can substitute for physician time and, if used wisely, can expand output by freeing physicians for tasks more directly suited to their training (Brown, 1988). Similar opportunities exist between physicians and nonphysician services in producing mental health services (Deb and Holmes, 1998). Expanded functions for dental assistants and dental hygienists have offered lower dental care prices without loss of quality (Liang and Ogur, 1987). Likewise, research suggests possibilities for substitution between registered pharmacists and pharmacy technicians (Okunade and Suraratdecha, 1998).

Nurse anesthetists are already generally accepted in practice, and while laws often limit what midwives can do, they have a long history of delivering babies. Finally, a stronger general emphasis on providing information to patients has made a wide variety of caregivers into information providers as well. The bottom line? Flexibility in the production of health care exists, in fact, and opportunities for substitution abound.

Assume that the decrease in the slopes (i.e., the ratio of factor inputs) is 6 percent. The elasticity of substitution, E_s, has a value of 0.6, indicating that every 1 percent change in relative factor prices leads to a 0.6 percent change in the relative use of those factor inputs. Whereas the ratio of physicians to nurses was previously 1.00, a 10 percent increase in relative physician wages (from 5 to 5.5 times nurses' wages) would change the input ratio to 0.94 (a 6 percent decrease). This would represent the replacement of one physician ($220,000) with five nurses ($200,000), hence lowering costs (due to the change in physician salary) by $20,000.[1] We will report E_s in absolute value as positive numbers, even though we understand that the firm's response is to decrease the relative use of a more expensive input.

The minimum value of E_s is zero, and a firm with isoquants represented by the one shown in panel A of Figure 6.1 will have an elasticity of substitution equal to zero because it always

Table 6.1 Substitution Elasticities for Teaching and Nonteaching Hospitals Evaluated at the Mean

Input Pair	*Nonteaching Case-Mix Adjusted Admissions*	*Teaching Case-Mix Adjusted Admissions*
1. Medical Staff with Nurses	0.547	0.159
2. Medical Staff with Beds	0.175	0.155
3. Nurses with Beds	0.124	0.211
4. Nurses with Residents	—	2.127
5. Medical Staff with Residents	—	0.292

Source: Gail A. Jensen and Michael A. Morrisey, "The Role of Physicians in Hospital Production," *The Review of Economics and Statistics* 68:3 (August 1986): 432–442. Copyright © 1986 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology, with permission.

will use the same input combination to produce a given level of output regardless of relative factor prices. Higher values of E_s indicate a greater potential for substitutability.

A good example of how elasticity of substitution is used is provided by Acemoglu and Finkelstein (2006). These authors observed how health care firms responded to the introduction of the Medicare Prospective Payment System, which they found to cause labor to become more expensive. True to the theory of the firm with its responsiveness to the market, they found the firms' response was to increase the capital/labor ratio.

Estimates for Hospital Care

Physicians are important to productivity of all hospital inputs. Pauly (1980) and Jensen and Morrisey (1986) incorporated this fact into their analysis of the elasticity of substitution between hospital inputs. They estimated a production function for hospital care and generated the isoquants and estimated the elasticities of substitution. Patterns of input use were observed in the process across a large number of hospitals.

Table 6.1 shows the estimates of elasticities of substitution between pairs of inputs. For example, a 1 percent increase in the price of medical staff relative to nurses would result in a 0.547 percent decrease in the ratio of medical staff to nurses.

How are we to judge the estimated degree of substitutability? Are these numbers large or small? The elasticities reported are at least sufficient to show that some substitutability exists between virtually all pairs of hospital inputs. In fact, the authors conclude, "all inputs in both teaching and nonteaching hospitals are substitutes for each other." The smallest values for substitution reported here are between beds and categories of labor. It may seem nonsensical that one can substitute people for beds, but "beds" here represents a convenient measure of the various and often complex capital inputs used by a hospital. The data suggest that even in hospital care production, where labor is undoubtedly the critical input, capital can be substituted on the margin for labor.

Custer and Willke (1991) and Lehner and Burgess (1995) have shown these results to be sensitive to the particular definition of the physician's input. Even so, most health economists probably would agree that substitution possibilities exist among many health care inputs, though their range is still uncertain.

Recent studies have shown that home health care can substitute for hospital care. Is the "home hospital" as good? Is it less costly? Illife and Shepperd (2002) report that studies of randomized trials of home versus hospital care find little difference in mortality outcomes or in costs.

Costs in Theory and Practice

The production function, its isoquants, and the elasticities of substitution have consequences for costs. We will show the derivation of the cost function and explain the technical terms *economies of scale* and *economies of scope*.

Deriving the Cost Function

As we noted, the production function describes the input/output relationships, and the cost function describes the cost/output relation. The two are closely related, and under the right conditions, the two functions can be derived one from another. We illustrate the closeness of this relationship in Figure 6.2.

Panel A, which depicts several isoquants, illustrates the production function, let's say for a physician practice. An infinite number of isoquants exist in principle, and we have chosen to show only three. The lowest isoquant shows that many points (or combinations) of capital and labor are capable of producing the 100 physician office visits. Higher isoquants, of course, produce more, 150 and 200 visits, respectively. Imagine that Figure 6.2, panel A, had no straight, slanted lines (*AB* and so on). The lines "removed" identify the cost consequences of the firm's input choices. Without them, the graph would characterize a firm that knows "only half of the story," only the output consequences of its input choices. Do firms exist that would simply ignore the cost inputs? Would they say: "We wish to help people get well regardless of what it costs and economics is not an appropriate consideration."

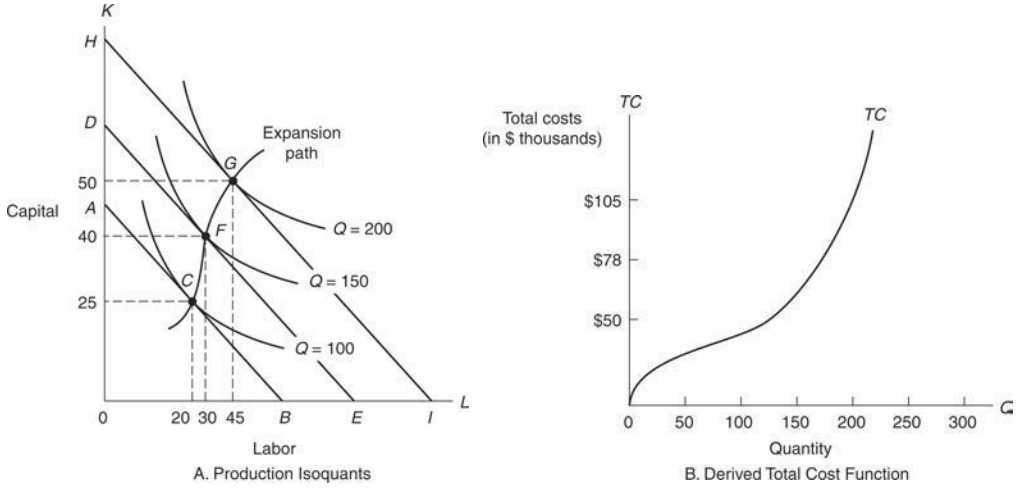


Figure 6.2 Production Function for a Hypothetical Physician Practice

In reality, health care firms, like all others, must meet their creditors or close their doors—they must at least break even. A tougher question for economists, however, is whether the many nonprofit firms strive to minimize costs. We must further ask, how well do they achieve this goal? Economic theory derives the cost function only if the firm in question seeks to minimize its costs. Before showing how the derivation works, consider that Ellis (1993) makes a good case that we are "safe" in this regard. Ellis addresses this question: If a community hospital board discovered that it could treat 1,000 patients with proper care and still permit a waste of resources worth $20,000, then some board member could sensibly suggest: "Eliminate the waste and we could treat even more patients." We will assume that health care firms reason in this way.

Cost Minimization

While isoquants show the many combinations of inputs to produce a given output, they do not by themselves describe the cheapest combination of inputs to produce the given output. The firm minimizes costs of producing a given level of output with the aid of a second tool—the isocost curve, the downward-sloping straight lines in Figure 6.2.

An isocost curve is the collection of all combinations of capital and labor that together cost a given amount. Let the level of total cost being considered be TC; this money can buy many combinations of capital and labor, K and L, in fact any combination whose costs add up to $TC = rK + wL$, where r is the rental price of capital and w is the wage rate of labor. The "rental price of capital" means the cost to the firm of using the capital for one period, regardless of whether it actually rents the capital or owns it. This particular isocost equation can be transformed algebraically so that K appears on the left-hand side and all other terms appear on the right, yielding the equivalent equation, $K = TC/r - (w/r)L$, which is an isocost line.

The firm wishing to produce a given output level, say 100 visits, will minimize its costs by choosing the lowest isocost curve that is tangent to the 100-visit isoquant. In the figure, least-cost production of 100 visits occurs at input combination $L = 20$, $K = 25$ at point C on isocost curve AB. Given knowledge of the input prices, we can calculate the output cost. For example, let r = $1,200 and w = $1,000; then, least-cost production of 100 visits will cost $50,000. Alternatively, if the firm wishes to produce 150 visits, the least-cost production would occur at point F, which entails 30 units of labor and 40 units of capital for a total cost of $78,000. In this fashion, the combination of the production function represented by the isoquants and the cost requirements represented by the isocost curves generates a set of outcomes, or points C, F, and G. The set of all possible points of tangency, such as these, is called the expansion path.

The expansion path supplies the information that associates a given output with its minimum cost. When these cost and output data are recorded in a graph, as in panel B, the result is the firm's total cost function. The cost function has a lazy S-shape, a pattern thought to be typical of many firms in practice. It also goes through the origin, indicating that if this firm produces nothing, it will incur no costs, meaning that the firm has no fixed costs. Economists refer to a period long enough for the firm to alter or avoid any of its commitments to input suppliers as "the long run." Thus, the implication is that the cost function shown is depicting the firm in the long run. This cost function is a "frontier" in that it represents the minimum possible cost of producing a given output. Actual firms may operate somewhat inefficiently, and we will observe cost levels above the frontier. It is a contradiction to say that one could observe a cost level below the cost frontier. Analysts often wish to determine whether health care firms are operating on or above their cost frontier, and we will return to this efficiency issue later. At present, we address a different issue—whether a firm is operating at an economical point on the frontier from the point of view of society as a whole.

Economies of Scale and Scope

To simplify the exposition, we separate the issues of economies of scale and scope, treating scale economies first. Consider a physician firm such as the one depicted in Figure 6.2. The long-run total cost function as shown in panel B can be transformed to express information about economies of scale. The average costs for this firm can be calculated by dividing the given cost level by the corresponding number of physician visits. The resulting long-run average cost (*LRAC*) function is in Figure 6.3. A firm experiences economies of scale when its long-run average cost is declining as output increases. Thus, the firm depicted exhibits economies of scale in region *AB*. Conversely, the firm experiences diseconomies of scale if and only if the long-run average cost is increasing as output increases, such as occurs in the region *BC*. What output level would a profit-maximizing firm choose in this case? It is tempting to suppose it would choose output Q_B, at which its average costs are lowest (AC_B), but this is not necessarily the case. A firm is not in the business of minimizing its average costs and would do so only if, coincidentally, the output that minimized costs also maximized profits.

A related concept is economies of scope. By definition, economies of scope are possible only for a multiproduct firm and because many health care firms are multiproduct in nature, the concept is highly relevant. Economies of scope occur whenever it is possible to produce jointly two or more goods more cheaply than if we produce them separately. As an illustration, consider the provision of pediatric hospital care (for children) and geriatric hospital care (for the elderly). Suppose there were two hospitals in town—one that provided only pediatric care and one that provided only geriatric care. Would the total cost of pediatric plus geriatric care be lower if one single hospital provided both? It may be cheaper to combine the two hospitals and achieve scale economies, but that is not the point at present. However, it might still be cheaper to combine them when the inputs needed for both types of care interact well together. Perhaps things learned in pediatrics have applications in geriatrics and perhaps the two could support each other so that the result would be lower total costs. If so, by producing the two different outputs jointly, we achieve economies of scope.

Economies of scope are illustrated by equation (6.1). In mathematical notation two outputs are shown, Q_1 and Q_2. For example, Q_1 may represent pediatric care and Q_2 may represent geriatric care. The example in equation (6.1) concerns the attempt to produce output levels $Q_1 = 100$ and $Q_2 = 150$ either jointly or separately. Economies of scope exist if the cost of producing the two outputs jointly, that is,

$$TC(Q_1 = 100, Q_2 = 150)$$

is less than the sum of the costs of producing each quantity separately, that is,

$$TC(Q_1 = 100, Q_2 = 0) + TC(Q_1 = 0, Q_2 = 150)$$

In summary, economies of scope occur in this example if the following inequality holds:

$$TC(Q_1 = 100, Q_2 = 150) < TC(Q_1 = 100, Q_2 = 0) + TC(Q_1 = 0, Q_2 = 150) \qquad (6.1)$$

Why Would Economies of Scale and Scope Be Important?

The concepts of economies of scale and scope are of considerable interest to both public policy and to managerial policy. We illustrate their significance using the idea of economies of scale, but we can use similar arguments to show the importance of economies of scope.

Recall that a profit-maximizing firm wishes to do just that, maximize profits. It has no intrinsic interest in producing at that level of output at which average costs are at a minimum

unless that output also coincidentally maximizes profit, but this is not generally the case. Consumers would like firms to minimize average costs as long as the firms pass these cost savings on to the public. One of the reasons that economists promote the theory of perfect competition is that competition forces the firm in the long run to operate so that it minimizes average costs. The competitive firm is guided by competition, as if by an invisible hand, to serve society's interests in keeping costs low.

Most health care firms, such as hospitals and physician practices, do not operate in perfectly competitive markets. Therefore, competitive pressures will not necessarily force them to operate at the most efficient scale of operation. There may conceivably be too few or too many providers. If there are too many, existing hospitals may be smaller than is required for the efficient scale; small hospitals may be forgoing the profit-maximizing opportunities available through greater economies of scale.

Historically, area-wide health planning, promoted by various government programs since the 1940s, encouraged the reduction of "excess beds." These programs also promoted the "rationalization" of geographic patterns of critical and expensive diagnostic equipment, such as CT scanners. The programs expressly aimed at reducing the growth in health expenditures would make more sense to economic thinking if the health care were either unnecessary or if the rationalization were expected to take advantage of economies of scale.

The gain to society from exploiting economies of scale also illustrates the natural monopoly. Consider a simple example. Assume that Figure 6.3 shows a health care unit (perhaps a hospital) that provides the diagnostic services of an MRI scanner. If many such firms are in the market area, then perhaps no single MRI scanner is operating at an output level at which it achieves the lowest long-run average cost. Because of its competition, the hospital unit depicted operates at point *A*, where average costs are higher. If society had fewer MRI scanners, the remaining ones could operate at a higher capacity, say at point *B*, with lower average costs of AC_B. These arguments depend on finding empirical results showing that further advantages from economies of scale at the level of the firm are available to society as a whole. The average cost curve in Figure 6.3 records only the costs incurred by the hospital unit. From society's broader perspective, the costs incurred by the patients and their visiting families and friends also are relevant. To see the point more clearly, draw the MRI scan

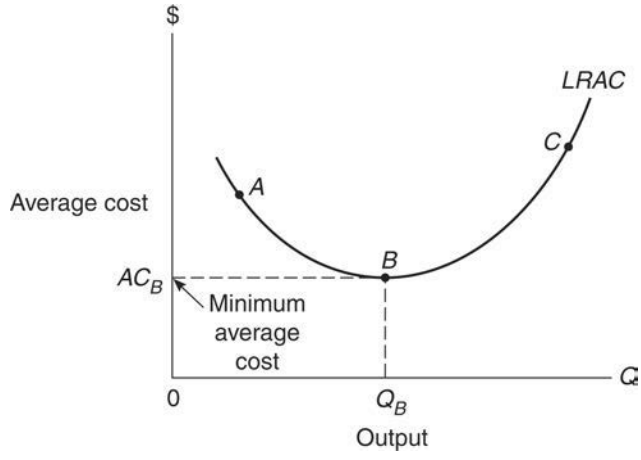


Figure 6.3 The Long-Run Average Cost Function

example to its extreme. Suppose the minimal average cost occurs when one scanner serves a rural region of 400 miles in radius. Would residents of this region be better off building only one centrally located scanner? Or would the necessarily large travel costs make such a plan foolish in the extreme?

Empirical Cost-Function Studies

With the theoretical ideas understood, the fundamental questions become empirical. Are there, in fact, economies of scale and scope available to be exploited in real-world health care firms? At what level of output and for what combinations of outputs are these economies achieved? Two themes occur often and represent differences in approach. We will describe long-run versus short-run studies, and behavioral versus structural cost functions.

LONG-RUN VERSUS SHORT-RUN STUDIES We have seen how the shape of the long-run average cost curve defines economies of scale. The difference between the long run and the short run is well-defined in economics. The long run is a period sufficiently long for the hospital to end any fixed commitments and to make any cost-saving adjustments that are possible. The short run is a period during which the hospital still has some fixed commitments, that is, some inputs that cannot be varied. An example is the number of beds set up for service. Research (Vita, 1990; Fournier and Mitchell, 1992) has demonstrated that results differ depending on the investigator's assumption of a long-run versus a short-run equilibrium.

Economists use short-run cost estimates to test for the short run or long run. Econometricians can assess this question by simulating that the capital is variable in the estimated function. If simulated profits rise when capital is changed, they know that the firm is operating in the short run. Bilodeau and colleagues (2002) recently estimated a cost function which found the U.S. hospital system to be overcapitalized, which we interpret as short-run phenomena.

STRUCTURAL VERSUS BEHAVIORAL COST FUNCTIONS Economic and health service analysts frequently distinguish between structural and behavioral cost functions. By structural cost function, we mean a cost function derived in a consistent manner from economic theory, just as we have derived it in the previous section. That is, we use the production isoquants and the isocost curves to derive the cost-minimizing level of costs for each possible level of output.

In contrast, behavioral cost functions (Evans, 1971) are derived from analyses of the patterns in costs in actual data across hospitals. Variables are included that distinguish real-world differences between hospitals. For example, teaching hospitals have higher costs due to the teaching and research services that they provide (Farsi and Philippini, 2008; Linna and Häkkinen, 2006). The variables matter for costs but often do not have a clear role in the theory of cost functions. Researchers have also the role of market structure on hospital costs, such as the degree of competition (Banerjee and Cohen-Cole, 2012; Jiang, Friedma, and Jiang 2013). Sometimes behavioral cost functions omit variables, like factor costs such as employee wage rates or equipment.

Of those who chose the theoretically most consistent "structural" approach, some (Conrad and Strauss, 1983) found economies of scale, some (Cowing and Holtmann, 1983) found constant returns to scale, yet others (Vita, 1990) reported diseconomies of scale. Researchers who applied the behavioral cost-function approach (Granneman, Brown, and Pauly, 1986) found economies of scale for the emergency department.

Difficulties Faced by All Hospital Cost Studies

What do hospitals produce? The difficulty of measuring the heterogeneous hospital output occurs in nearly all hospital cost studies. First, hospitals differ by type of cases they treat; this is

the "case-mix problem." Medicare's Diagnosis Related Group (DRG) payment system identifies 745 groups of cases, so the hospital is a multiproduct firm to an unusual degree. Some studies have virtually neglected the problem; others have used multiproduct cost functions with adjustments for case mix and related variables. Although the multiproduct approach is superior, even multiproduct methods rarely incorporate more than four or five hospital output categories. Even the question of whether sufficiently refined hospital cost estimates are achievable in principle has not gained a consensus among health economists.[2]

A related problem is how to treat quality. Unobserved or incorrectly measured variations in quality between hospitals may lead to errors in research examining economies or diseconomies of scale. Recent research established that hospital quality affects hospital costs, and quality is more expensive (Carey and Stefos, 2011).

The case-mix issue is illustrated in Figure 6.4, and the quality issue, though not illustrated explicitly in the figure, is similar. The figure illustrates a situation where the true, long-run average cost function is flat, exhibiting neither economies nor diseconomies of scale. The three cost curves shown represent three hospitals, each with a distinct case mix. In order of ascending costs, these are Hospital 1, which treats uncomplicated medical cases; Hospital 2, which treats more complicated surgical cases; and Hospital 3, which treats the most highly complex level or tertiary cases.

Suppose that points *C*, *D*, and *E* represent the data observations available to the researcher for each hospital type for a given statistical study. The researcher may mistake the unobserved case-mix differences for diseconomies of scale, that is, mistakenly believe that the long-run average cost curve is rising as shown by the connecting line. This case is illustrative only because the underlying patterns of case mix, quality, and size of output could yield errors in either direction. Alternative approaches to treating the case-mix problem have been investigated extensively.[3] These studies show that case-mix differences between hospitals are materially important when estimating scale economies and cannot be overlooked.

Researchers may also lack reliable measures of hospital input prices. The cost of capital or the starting registered nurse's wage may seem simple data to collect (a matter of contacting an executive in the finance office), but they are often difficult to define adequately. Unfortunately, errors in measurement of hospital input prices have substantial consequences for the results.

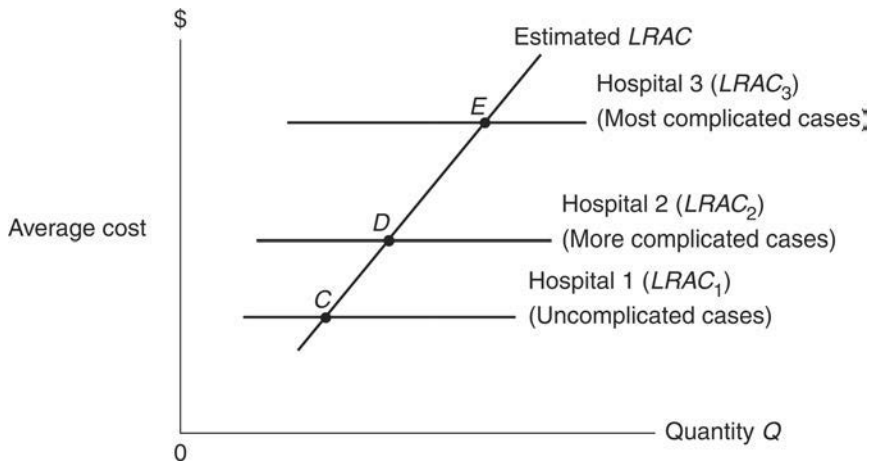


Figure 6.4 The Case-Mix Issue

Finally, hospital cost-function studies almost always omit physicians' input prices entirely. Physicians are not generally under hospital control, and this independence, as well as the accounting and the physician practices of billing the patient separately, become problems. Physician costs are properly part of production costs, and their omission results in biased econometric estimates of hospital costs. Early research experience with hospital costs (Pauly, 1980) established that physicians do matter. These studies establish that the difficulties in hospital costs center on the problems of measuring output in a multiproduct firm where quality matters a great deal.

Modern Results

The most recent research supports claims that economies of scale exist in hospitals. Preya and Pink (2006) studied costs of Canadian hospitals prior to a massive consolidation, finding "large scale unexploited gains to strategic consolidation in the hospital sector" (p. 1049). Dranove and Lindrooth (2003) studied a large number of hospital consolidations, comparing them to matching hospitals that did not consolidate. They found "significant, robust, and persistent savings for mergers, 2, 3, and 4 years after consolidation" (p. 996).

Summary: Empirical Cost Studies and Economies of Scale

Early hospital cost studies led economists to believe that economies of scale existed, even claiming that the optimal hospital size was about 250 beds available for patient care. However, there followed a flurry of criticisms and corrections; the complex multiproduct hospitals, which also varied in quality, required more sophisticated methods. Studies that followed tended to dispel the earlier consensus with widely varied and sometimes contradictory results. Modern work (studies appearing since 2000) much more clearly reports that economies of scale exist in hospitals, a result that suggests that many hospital mergers might be justified on the basis of cost savings to society. Economies of scale are now more often studied for specific hospital services (Goncalves and Barros, 2013).

Technical and Allocative Inefficiency

In addition to issues of scale, efficiency can be measured in two other ways, each also of great concern to health care firms and policymakers. Economists refer to these as technical and allocative efficiency, or their lack, which is inefficiency.

Technical Inefficiency

Technical inefficiency is illustrated in Figure 6.5, panels A and B. Panel A depicts a production process with one input, while panel B depicts a production process using two inputs, capital and labor. Technical inefficiency implies that the producer is not achieving a maximum output from a given input combination. It is as if workers or machines were misused, not working at full capacity, or not cooperating well. In both panels, each firm's actual experience is indicated by a firm number. A technically inefficient firm falls off its frontier. In panel A, the production frontier is shown as the production function labeled *f(L)*. Firms 4 and 5 are on the frontier; that is, they are currently technically efficient. In contrast, Firms 1, 2, and 3 are off the frontier and thus are currently technically inefficient. In panel A, the firm inefficiencies

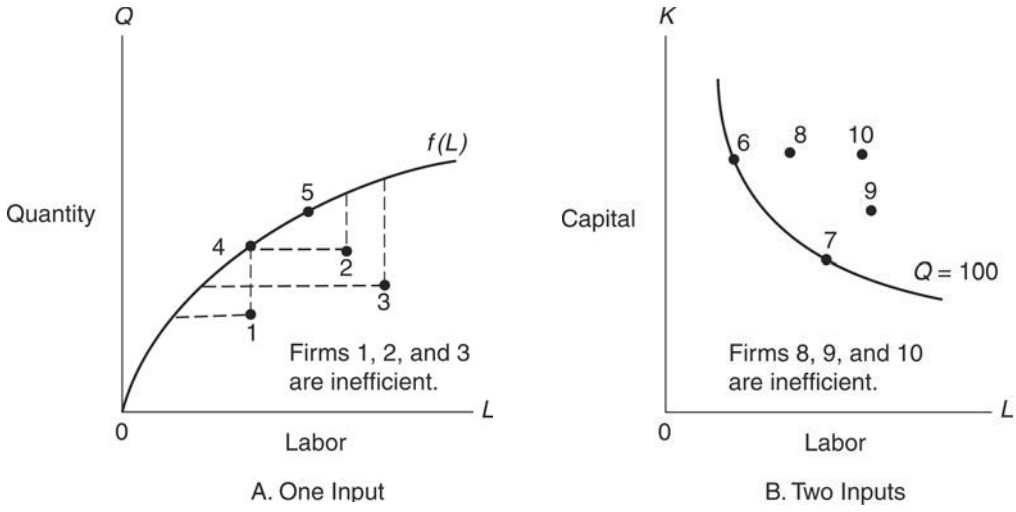


Figure 6.5 Technical Efficiency and Inefficiency at the Firm Level

are measured as relative distances from the frontier. The distance may be the output (vertical) distance—what output could have been achieved with these inputs—or it may be the input (horizontal) distance—how many fewer inputs could have achieved this output? Generally, these alternative approaches to inefficiency measurement will yield somewhat different results.

Panel B shows an isoquant representing frontier practice treating for 100 cases. Suppose that the actual current output of all firms depicted in panel B is known to be 100 cases. Firms 6 and 7 in the panel are on the isoquant for 100 cases and thus currently represent frontier practice and are technically efficient. Firms 8, 9, and 10 are off the isoquant, indicating that they have employed more input quantities than technically efficient production requires. As was the case in panel A, both output and input distance functions can be used as measures of inefficiency.

Allocative Inefficiency

Technical efficiency applies conceptually to production within a given firm. By contrast, allocative efficiency requires the efficient allocation of inputs between firms and between outputs. Essentially, it requires that each type of capital and labor be put to its most rewarding use in society. Economic theorists have shown that allocative efficiency in production will result if each firm buys or hires inputs in competitive markets and if each firm minimizes production costs. Assuming competitive input markets—and thus fixed input prices common to all firms—permits us to describe allocative efficiency at the firm level. Here, allocative efficiency requires that each firm respond optimally to input prices; correspondingly, allocative inefficiency implies choosing an inappropriate combination of inputs in the sense that inputs and their prices have not been appropriately considered.

To illustrate, consider Figure 6.6. In this figure, the isoquant for 100 cases is illustrated as the curve labeled $Q = 100$. Assume that the firm being examined is currently producing its desired level of output, and that the desired level happens to be 100 cases. Two isocost curves

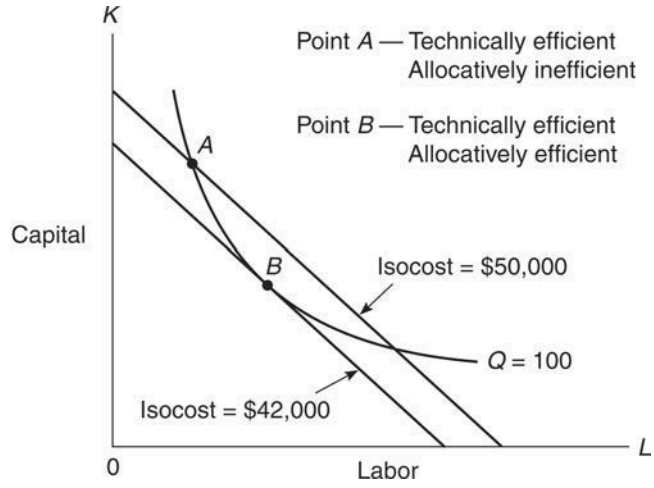


Figure 6.6 Technical Allocative Efficiency

also are depicted, with one indicating a cost level of $50,000 and the other a cost level of $42,000. Suppose that the firm in question was observed operating at a point *A*. Because this firm is treating 100 cases using an input combination on the 100 cases isoquant, we can say this firm is technically efficient. However, it is not allocatively efficient. At the current input prices, it uses too much capital, and not enough labor.

To confirm this idea, consider that the firm in question alternatively could have produced 100 cases at point *B*. As point *B* lies on a lower and thus less costly isocost curve, the firm at *B* would reduce costs from $50,000 to $42,000 by moving from point *A* to point *B*. Point *B* entails a tangency of the desired isoquant with the lowest feasible isocost curve. A tangency implies an equality of the ratio of input prices to the ratio of marginal products for the inputs (the equality of marginal output per dollar for each input). This is the firm's appropriate response to input prices, and its key condition for allocative efficiency.

Though the conditions for efficiency of both types are well-defined, several different empirical techniques have appeared to address them. These techniques can be grouped into two categories: nonfrontier and frontier studies. In the nonfrontier studies, actual outputs or cost experiences for two or more groups of firms are compared while attempting to control for the effect of extraneous variables. In frontier studies, actual outputs or firm costs are compared to the best possible experience. We emphasize frontier studies because they are conceptually closer to the definitions of technical and allocative efficiency.

Two types of empirical frontier analysis have emerged. One, data envelopment analysis (DEA), was developed earlier. The other, the stochastic frontier analysis (SFA), was developed by 1977, with applications in health care proliferating in the late 1980s.

Frontier Analysis

DATA ENVELOPMENT ANALYSIS (DEA) The DEA approach is illustrated in Figure 6.7. The frontier production, initially unknown to the researcher, is revealed as more firms are observed. A few such firms are represented by the points labeled 1 through 10.

The data envelopment method finds the frontier isoquant for a selected level of output (for example, $Q = 100$) by forming an envelope of the data. Researchers use linear programming to construct this efficient outer shell of the data points. Given the estimated frontier, technical inefficiency then is measured as a relative distance from the frontier. DEA is particularly useful for hospitals in that it easily handles multiple inputs and multiple outputs.[4]

DEA attracts researchers and research readers because it imposes no assumptions about the parameters of the underlying distribution of inefficiency. Analysts refer to this as "nonparametric." Its cousin SFA contrasts with DEA because researchers must guess the statistical distribution of the inefficiencies in advance. Those who prefer SFA argue that DEA assumes that all firms lying distant from the frontier are inefficient. Suppose that during a given year, the Huron Hospital nurses go on strike. Suppose further that the hospital is measured as experiencing substantial inefficiency; the hospital was operating at a substantial distance off the frontier for similar hospitals. Was the hospital really inefficient or only apparently so? Should the researcher charge the entire shortfall in output to management error or should an adjustment be made (a "handicap" frontier) reflecting its special difficulties?

STOCHASTIC FRONTIER ANALYSIS (SFA) The SFA approach treats each firm uniquely by assuming it could be affected by a potential shock to its ability to produce care. Consider a hospital that is managed with perfect technical and allocative efficiency. Suppose this hospital discovers that its long-time major supplier has gone bankrupt. It takes months before comparable prices, qualities, and reliability are restored with new suppliers. Regardless of management's response to these events, the hospital's cost and output data for the year will differ from other apparently similar hospitals having standard experiences. This will be true even if the management is "perfect," a model for other managers.

If each firm is randomly shocked during the period in ways that affect its production and cost performance, the firm's best possible practice, its frontier, will be randomly shifted. When the frontier function is partly random, the result is a "stochastic process"—hence, the name stochastic frontier.

Techniques developed by Aigner, Lovell, and Schmidt (1977) make it possible to estimate both the individual firm's expected frontier and the firm's inefficiency. The stochastic

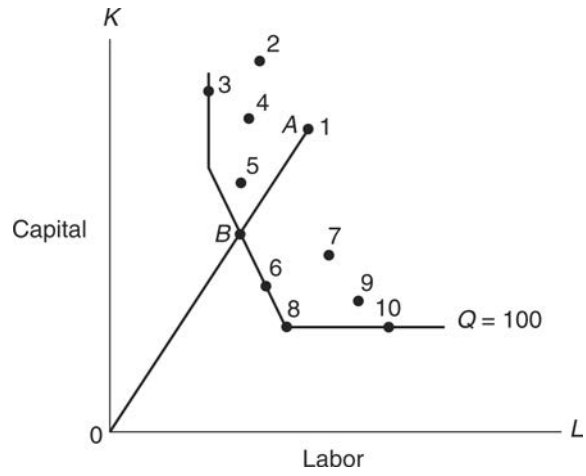


Figure 6.7 The Data Envelopment Method

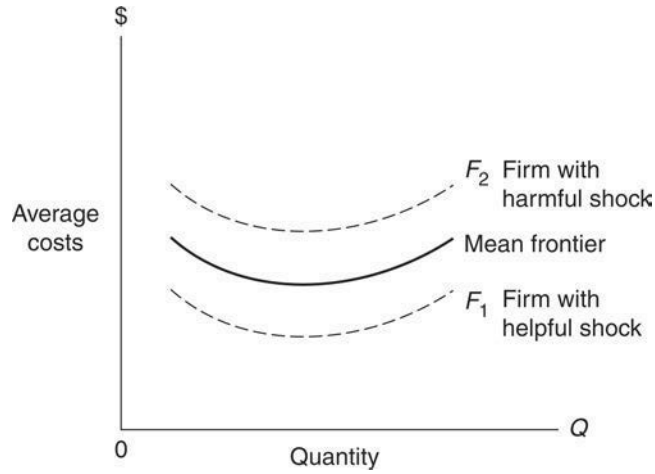


Figure 6.8 The Stochastic Frontier Method

frontier concept is illustrated in Figure 6.8, an example that focuses on average costs. The mean average cost frontier is the dark curve. The individual firm's frontier is shifted by random shocks whose distribution must be assumed by the researcher from among known, parametric distributions. Thus, each hospital has a unique frontier and inefficiency distance from its frontier.

Many health economists recognize pluses and minuses to both approaches and treat them as complementary tools (Kooreman, 1994a, 1994b). For example, Matawie and Assaf (2010) apply both methods in comparison. Work with "panel data"—gathered as a sample of hospitals followed for many periods—suggests how the SFA method can be used without imposing the strong parameter assumptions (Lee and Schmidt, 1993).

The Uses of Hospital Efficiency Studies

The estimates from the frontier analyses have stimulated investigations of substantial consequence. Consider these cases with background provided to explain the significance.

TOTAL HOSPITAL EFFICIENCY Critics argue that the U.S. hospital system has done an inadequate job of improving U.S. health status to justify its huge cost. We have seen that costs depend in part on whether one achieves economies of scale, and of course total costs also depend on the quantity demanded. But we have just seen that technical and allocative inefficiency also play a role. How efficient are U.S. hospitals?

Their reported efficiency levels have been quite high. The earliest DEA study (Valdmanis, 1990) reported technical efficiency levels of about 90 percent, while Magnusson's DEA (1996) study reached similarly high levels. SFA studies have tended toward similar levels; early SFA studies (Zuckerman, Hadley, and Iezzoni, 1994; and Folland and Hofler, 2001) found the sum of technical and allocative inefficiency to be only a little more than 10 percent. Since then studies have reported roughly similar levels of inefficiency. Frontier researchers are well aware of the many sensitivities of the method to variations across individual studies, but the various results support the hypothesis that the hospitals are highly efficient. One cannot blame hospital inefficiency for the high level of costs.

For-Profit versus Nonprofit Hospitals

Many people place greater trust in nonprofit hospitals than for-profits. Yet many economists and legal theorists are critical of nonprofit hospitals and demand to know why they deserve special benefits like freedom from taxes. To date, efficiency studies most often have not favored one organizational form over the other.

In many recent studies, nonprofit and for-profit hospitals appear approximately equal in efficiency. While the earliest studies (Valdmanis, 1990; Ozcan et al., 1992) found differences between samples of public and for-profit hospitals, studies since then found no significant differences (Sloan et al., 2001). Burgess and Wilson (1998, p. 100) found "no evidence that differences in ownership affect technical efficiency after controlling for other factors." Looking at Italian hospitals using DEA, Barbetta, Turati, and Zago (2007) show a convergence of mean efficiency scores between not-for-profit and public hospitals. They believe that differences in economic performances between competing ownership forms result more from the institutional settings in which they operate than the effect of the differing incentive structures.

If the efficiency data are neutral between organization types, this tends to favor the argument that nonprofits lack a justification for their tax breaks, at least on the grounds of technical and allocative efficiency.[5] Perhaps our impressions of relative efficiency of hospitals will change. Hollingsworth (2008) conducted a substantial literature review of over 300 frontier efficiency studies. Though cautious, he concluded that the public providers were somewhat more efficient.

Efficiency and Hospital Quality

Mary Deily and Niccie McKay (2006) explain that hospital inefficiency may reduce the quality of care. Both care inputs in their study have been adjusted for quality. Quality of hospital output is measured by mortality rates. In principle, the process of combining inputs together may be affected by inefficiency. These authors test the proposition in a sample of about 140 Florida hospitals measured over three years. They found, using the stochastic frontier approach, that the inefficiency measure was a highly significant and positive contributor to a measure of hospital mortality rates.

Laine and colleagues (2005) attempted similar tests for long-term care. Although they detected no inefficiency effect on "clinical quality," they found inefficiency to contribute to the prevalence of pressure ulcers, "indicating poor quality of care was associated with technical inefficiency" (p. 245).

Are Hospital Frontier Efficiency Studies Reliable?

This question was asked by both Folland and Hofler (2001) and by Street (2003). As an example, one of these papers estimated hospital efficiency values by three different versions of the stochastic frontier method. The versions were minor differences in the assumptions most investigators consider. Then they found the correlations between the versions. The correlations were rather poor, at 0.70 or lower. A rule of thumb here is that when the object is to test whether two series of numbers are valid equivalent measures of each other the correlation coefficient should be 0.70 or higher. The authors concluded that the frontier estimates seem adequate to discern mean differences between groups of hospitals, but they cannot be justified for the task of identifying inefficiencies by individual hospitals.

BOX 6.2

Should We Close Inefficient Hospitals?

Hospitals that are technically and/or allocatively inefficient will have higher costs than their more efficient peers. Should they be closed to save money? What about the utility loss of their former patients? Capps, Dranove, and Lindrooth (2010) provide an interesting way to answer the questions. On the one hand, the cost savings from closing a particular hospital will be partly offset by increases at other hospitals that pick up these patients. On the other hand, recall that since the peer hospitals are more efficient, they will have higher occupancy rates after the change.

The authors' method allowed them to calculate the total travel time in the market that would be equivalent to the utility loss. Using industry estimates of the opportunity costs of driving, they find the dollar equivalent of the utility loss. The authors applied their approach to the cases of five recently closed hospitals, and they found the hospital closings provided a net welfare gain for the system.

Performance-Based Budgeting

Yaisawarng and Burgess (2006) report success in devising an application of hospital efficiency data to the financial reimbursement methods of the Veterans' Administration hospital system. They have made a preliminary application to the hospitals, a method of payment to each hospital group in which the more efficient groups receive the highest payment. This provides healthier incentives, they argue, in contrast to the previous system where higher-cost hospitals received higher payment. If this works, it realizes a common dream for efficiency data, though it needs to overcome the skepticism generated by earlier research that found rankings of hospitals by efficiency scores to be sensitive to variations in estimation methods.

Technological Changes and Costs

The rapid pace of technological change in the health care industry raises economic questions about the effects these changes will have. Technological change may reduce costs when it improves the productivity of health care resources, or it may increase costs when it improves the quality of care or introduces new and costlier products. Because it often raises costs in the health sector, many researchers hypothesize technological change to be the major contributor to health sector inflation. Zivin and Neidel (2010) also point out that some technologies are irreversible and adoption for the patient may preclude some future treatments, such as some treatments with antibiotics that may lead to resistant strains.

Deep and widespread insurance coverage in the health sector may induce technological innovations of the type that increase costs. The effects on costs and the improvements to quality of care will depend on the diffusion of these new technologies to providers. Thus, patterns of diffusion have also become a critical subject of study. We investigate these issues in the remaining sections of this chapter.

Technological Change: Cost Increasing or Decreasing?

Technological change necessarily entails an improvement either by providing less costly production methods for standard "old" products, or alternatively by providing new or improved products. In either case, it will be less expensive to produce a given output, holding quality constant. However, the mix of products and services sold may change in directions that raise the average cost of a patient day, a case treated, or a physician visit. Consequently, the total health care expenditure per capita may rise.

Figure 6.9 illustrates these concepts. In panel A, we hold quality of care constant and illustrate the isoquants representing 100 cases before and after a technological change. The technological improvement in panel A shifts the isoquant inward. The firm chooses an efficient combination of inputs at point E and after the change at point E'. This change results in the attainment of a lower isocost curve for treating the 100 cases.

Panel B illustrates the introduction of a new technology that makes it possible to treat 100 cases with better health status outcomes, thus providing a higher quality of care. Improvements entailing new products or, as shown in panel B, improved quality of care, are beneficial to the consumer, but they may be more costly. This is illustrated by a shift outward of the 100-case isoquant, resulting in production on a higher and costlier isocost curve. The typical patient will pay more for care. In some cases where patients are heavily insured, we may question whether the change is worth it to patients when they pay increased insurance premiums, or to society as a whole.

Health Care Price Increases When Technological Change Occurs

How do we measure the cost of a treatment when the treatment changes radically over a mere one or two decades? For example, heart attack treatment (myocardial infarction) changed

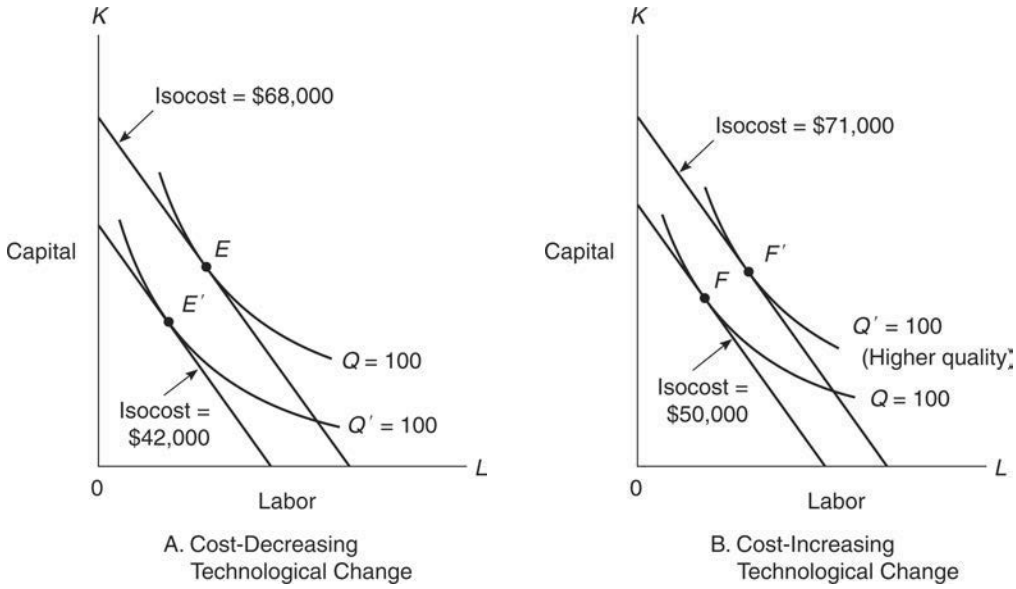


Figure 6.9 Cost-Decreasing and Cost-Increasing Technological Change

substantially from 1975 to 1995. Some new effective inputs proved extremely inexpensive (see the feature on aspirin in Box 6.3). Some materials did not exist in 1975, such as the intraortic balloon pump. Treatment practices changed; the average length of a hospital stay is now much shorter. Most important to the patient, the treatments are now more effective and have improved the length and quality of life for heart attack victims.

Treatment effectiveness improved, and in some cases less expensive inputs become available. Sometimes newly designed inputs were more costly than the ones they replaced. To see whether heart attack treatment in 1995 was more expensive per episode than in 1975, we must hold quality constant within the analysis.

BOX 6.3

Aspirin, the Wonder Drug at a Bargain

References to prototypes of aspirin date back to the works of the Greek physician Galen. These references mentioned salicylate-containing plants, such as willow bark and wintergreen. We today usually attribute aspirin to the Bayer Company in Germany in the later 1800s. Throughout its history, professionals praised aspirin for its excellent powers to relieve pain and fever (Andermann, 1996).

We think of this humble product as a cheap, over-the-counter drug that is widely available at a few pennies per dose. While fairly safe, it can have serious side effects with overdosing. Physicians, for many decades, said that "if aspirin had been proposed now as a new drug it would probably require a prescription." While it is doubtless that it will remain an over-the-counter drug, it now seems like a new product. Because of new discoveries of its benefits to heart patients, its influence continues to expand.

This is reflected in various treatment regimens on heart outcomes. David Cutler found that the basic three regimens—intensive technologies, non-acute pharmaceuticals, and behavioral change—have approximately equal contributions to improved outcomes. The non-acute drugs (pharmaceuticals) include those to control hypertension, reduce cholesterol, treat pain, dissolve clotting, and thin the blood. Aspirin is an effective blood thinner and providers now regularly prescribe it as a preventative measure as well as to reduce the recurrence of heart attacks (Cutler, 2001).

These are the essential problems of any price index. Consumers are familiar with the Consumer Price Index (CPI) which is used to measure inflation. A subset of this index focuses on medical care. In this case, the multiproduct character of hospital care can be confusing.

Until recently, hospital prices comprised selected components of hospital care. Room charge, nursing, lab, and other service centers each had a price index and then the results were combined into a hospital price index. Though useful, this approach often distorted the cost picture. For many years the room charge was weighted heavily, and as room charges soared, the index tended to exaggerate hospital price inflation. Even more confounding, hospital average length of stay declined rapidly in the United States, and the old price index neglected this saving. Although patients paid much more per day, it was offset in part by

shorter stays. Political debate centered on these price indexes, causing considerable misunderstanding. The DRG legislation that Medicare installed in 1983 reflected in part the wide public concern that health care inflation was out of control.

A good example of making the needed adjustment is a study done by Cutler and colleagues (1999). They developed two myocardial infarction treatment price indexes that measure patient gains in quality-adjusted life-years (QALYs). They evaluated these gains in dollar terms, and allowing for errors, particularly in valuing life-years, these authors proposed a range of price inflation estimates. Their research shows clearly that

1 Technological change makes a difference in patient lives.
2 The fact that substantial quality improvement had previously been omitted meant that previous estimates of inflation in health care needed to be reduced.

How well are their ideas corroborated by the data? Table 6.2 indicates that technological change has improved the quality of heart attack treatment and that this quality adjustment can turn what at first looks like price inflation into one of price deflation. Their further research suggests that the true decline in the heart attack treatment price might be much larger than these figures.

Table 6.2 Comparing the Unadjusted Approach with a Quality-Adjusted Measure of Price Increases of Treatment of Myocardial Infarction (1983–1994)

Index	*Avg. Annual Price Changes (%)*
Unadjusted Indexes	
Official medical care CPI	3.4
• Hospital component	6.2
• Room	6.0
• Other inpatient services	5.7
Heart attack unadjusted episode approach[a]	2.8
Quality-Adjusted Indexes	
Quality (extra years of life)	–1.5
Quality (extra QALYs)[b]	–1.7

Notes: [a] Experts recognize that several alternatives are applied when selecting for analysis the market basket of goods and services whose inflation is to be measured. The table reports a fixed-basket method; the patient is assumed to purchase essentially the same combination of medical goods in each year studied. Chain indexes allow for the representative market basket to change over time, and therefore add realism. But, how frequently should the basket be recalibrated? When a 6-year calculation of the basket is used, the average annual percentage change becomes 2.1 percent; with an annual recalculation, it becomes 0.7 percent.

[b] QALYs are quality-adjusted life-years. This quality of treatment is the same as the previous one except for the additional consideration of the degree to which the patient is able to lead a full, active life in the years after treatment.

Source: Adapted from Cutler et al. (2001). Reprinted with permission of University of Chicago Press.

Diffusion of New Health Care Technologies

It takes time for a new product innovation to be widely adopted by providers. Some firms adopt rapidly, some slowly, and some not at all.

Who Adopts and Why?

Those who study health technologies have found at least two basic principles that guide adopters: the profit principle and information channels. The first posits that physicians, for example, tend to adopt a new surgical technique if they expect to increase their revenues—this could happen through enhancing their prestige or by improving the well-being of their patients. The second is a compatible principle deriving originally from sociology, and it emphasizes the role of friends, colleagues, journals, and conferences in informing and encouraging the adoption decision.

Escarce (1996) emphasized the "information externalities" inherent in adoption by the first physician to adopt. An externality is the uncompensated, beneficial effect on a third person caused by the actions of a market, in this case, the actions of the first adopter. By adopting a technology, the physician communicates to friends and colleagues the expectation that the new product will benefit his or her patients and practice. The adopter paves the way for new infrastructure, new seminars, and library materials that reduce the cost of adoption for colleagues. The process tends to build on itself, perhaps at an increasing rate, until all the main body of potential adopters has acted, only then slowing the increase in total adopters until the community's maximum potential is reached.

The data are consistent with this process; in many industries, adoption occurs slowly at first, then at an increasing rate that continues at a decreasing rate asymptotically approaching its limit. Compare this description in words with Figure 6.10, which illustrates the classic pattern of diffusion as a logistic curve. A new data set is tested empirically by estimating the logistic function where P_t is the proportion of individuals or firms adopting by time t, where the maximum potential proportion of adopters is K, with parameters a and b to be estimated:

$$P_t = \frac{K}{1 + e^{-(a+bt)}} \tag{6.2}$$

Escarce's data fit this time pattern quite well. He studied a new surgical procedure, laparoscopic cholecystectomy, introduced in 1989, which is a minimally invasive technique to remove diseased gall bladders. He found the diffusion curve to fit the logistic pattern common to diffusion studies. He then examined the differences between those who adopted and those who did not. Adopting surgeons were more likely to be younger, male, board-certified, U.S. medical school graduates, and urban-located. Younger individuals are more likely adopters. In fact, in one study older physicians proved less likely to adopt even though their expected gains in profit were much higher than for their younger counterparts (Rizzo and Zeckhauser, 1992). Others suggest that followers are more likely to emulate the "star" physicians, ones with the most impressive credentials (Burke et al., 2007).

Other Factors That May Affect Adoption Rates

Economists believe that a firm will tend to adopt an innovation when the present value of future profits due to the innovation is positive. Waiting too long may provide competitors with an advantageous share of the market, which may be permanently sustained. However, waiting has benefits in that one may take advantage of future advances and learn from the

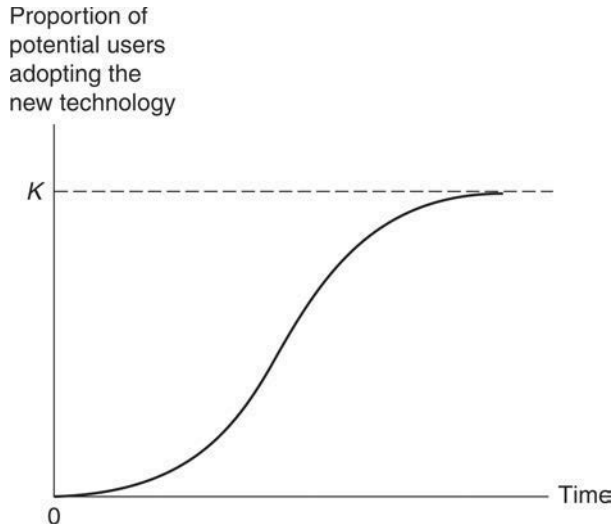


Figure 6.10 The Diffusion of a New Technology

experience of others. Waiting may reduce risks so more risk-averse firms may choose to wait somewhat longer. (For an introduction to the literature on hospital adoption of information systems, see McCullough, 2008.)

Sloan and colleagues (1986) found that mandatory rate-setting programs retard diffusion of technology in some cases, particularly in the instances of coronary bypass surgery, morbid obesity surgery, and intraocular lens implants. However, the degree of slowing tended to be small. Teplensky et al. (1995) conclude that restrictive rate-setting programs tend to retard significantly the adoption of new technology. Also, Caudill and colleagues (1995) report a slowing of the adoption of new blood dialysis technology when faced with restrictions on health care capital investment imposed by Certificate-of-Need (CON) legislation. CON laws, which are generally applied at the state level, require hospitals to gain approval from planning agencies when they wish to expand or to make a major capital purchase.

Diffusion of Technology and Managed Care

Managed care arrived with hopes that it would control health care expenditure increases by removing the financial incentives for physicians to overprescribe, overtreat, and overhospitalize their patients. The same flattening of incentives—no extra money for extra treatment—potentially dampens the physician's interest in cost-increasing technological change. As we have seen throughout this chapter, the reduction in incentives must be expected to slow innovation and the adoption of innovative technologies. Although we address the ultimate effect of managed care on health care inflation elsewhere in this text, it is appropriate here to ask: "Does a higher penetration of managed care into the health system tend to slow the growth in availability of new technologies?"

The answers seem to be "yes" for some technologies and "no" for others. Baker (2001) compared penetration of HMOs with adoptions of a new technology, magnetic resonance

imaging (MRI); he found a slowing of adoptions associated with HMOs. Baker and Phibbs (2000) found that greater HMO penetration also retarded the adoption of neonate intensive care. Hill and Wolfe (1997) examine a managed care-like system in Wisconsin. These authors reported time trends of adoptions for selected technologies, and the data suggest a retardation of several of the technologies but continued growth of several others. Friedman and Steiner (1999) investigated the availability of intensive care units and found no difference in admission rates under managed care versus fee-for-service care.

Conclusions

This chapter examined issues on the supply side of health care. Health care is fundamentally a production process, and it shares many characteristics with economic production generally. The production function, which summarizes the relationship of inputs and outputs, also embodies the technology. Technology that permits substitution between inputs provides better flexibility to the manager. The neoclassical cost function derives from the theory of production together with the theory of profit-maximizing behavior. Cost estimation describes the cost curves, which identify the economies of scale and scope. Health care firms may fail to achieve allocative or technical efficiency, or both. These analyses search for differences between for-profit, not-for-profit, and other kinds of firms.

Health firms may differ in technology because the adoption of new technologies differs among firms and is never instantaneous. Technology improvements in health care production may either increase or decrease costs depending on their effect on quality. Both market structure and regulation can affect the speed at which innovations are adopted.

Summary

1 Health care goods and services can frequently be produced in different ways in the sense that they use different combinations of factor inputs.
2 The elasticity of substitution is used to measure substitution. It represents the percentage change in the ratio of factor inputs resulting from a 1 percent change in relative factor prices.
3 Economists have found some substitution not only among different kinds of medical staff but even between hospital beds and medical staff, as well as with the application of the large variety of allied health professionals.
4 The principles of cost minimization, as represented by the locus of tangencies between the firm's isoquants and isocost curves, are used to derive the cost curves (total and average). Economies of scale refer to a declining long-run average cost. Economies of scope represent situations where the cost of producing goods jointly is less than the sum of the costs of separate production.
5 Early empirical work on hospitals found evidence of economies of scale and an optimum size of about 250 beds. Several recent contributions find economies of scale that depend on the nature of the hospital.
6 Technical inefficiency occurs when a firm fails to achieve the maximum potential output from a given set of inputs. It can be measured as a relative distance from the frontier production function or correspondingly as a distance from the isoquants of the frontier production function. Allocative inefficiency arises in the case of competitive input markets when a firm fails to purchase inputs, given their prices, in a manner that minimizes costs.

7 We measure efficiency by frontier methods of two types. The data envelopment method estimates the frontier by statistically forming an envelope of data points representing production data. The stochastic frontier method estimates simultaneously a randomly shifting frontier and the firm's distance from that frontier.
8 Technological change in health care may reduce costs if it improves production technology of existing products or increase costs if it emphasizes new products and higher-quality versions of old products.
9 New health care technologies are adopted gradually in an industry, and the pattern of adoption fits the logistic curve. Adoption patterns are influenced by regulation, age of the adopter, profitability, and channels of communication.

Discussion Questions

1 Explain whether there is any difference between goals in maximizing output for a given cost or minimizing the cost of producing a given level of output.
2 What are cross-sectional data? Why do economists find it so critical to control for case mix in studying health care cost functions? What are the analytical dangers if they do not?
3 How do legal restrictions on practice for nurses and physicians tend to affect the observed elasticities of substitution? Would the elasticities be higher if legal restrictions were removed? Would quality of care be affected?
4 Given the cost function and economies of scale and scope information reviewed in this chapter, does a policy encouraging large, centralized hospitals seem wise? Will market forces tend to reward centralization of hospital services?
5 Speculate on what types of services are more appropriate to large, regional hospitals, and what types of services are more appropriate to small, local hospitals.
6 Economists define the elasticity of substitution as the percentage change in the capital/labor ratio elicited by a 1 percent change in the factor price (wages/capital costs, for example) ratio. Would you expect the elasticity of substitution to be positive or negative? What would be the elasticity of substitution of a set of right-angled isoquants? Why?
7 Contrast technical and allocative efficiency. How can technical and allocative inefficiency in health care firms affect patient welfare?
8 What does "stochastic" mean in stochastic frontier efficiency estimation? Give several real-life examples of events that could shift the production frontier.
9 Which of the following types of technological change in health care are likely to be cost increasing: (a) threats of malpractice suits that cause physicians to order more diagnostic tests on average for a given set of patient symptoms; (b) a new computer-assisted scanning device that enables physicians to take much more detailed pictures of the brain; (c) the introduction of penicillin earlier in this century; (d) greater emphasis on preventive care? Discuss.
10 As technologies diffuse, why do some firms adopt them before others? What types of technologies would you expect to be adopted most quickly? Most slowly? What factors can slow the rate of diffusion of new medical technologies?

Exercises

1 Draw an isoquant that shows relatively little substitution between two factor inputs and one that shows relatively large substitution. Let the vertical axis represent capital and let the horizontal axis represent labor.

2 Draw isocost curves that are tangent to your isoquants in Exercise 1 and that each have the same slope. Mark the points of tangency and note the capital/labor ratio. Draw new, flatter isocost curves that are tangent, again each having the same new slope. Mark the points of tangency and note the capital/labor ratio. In which case is the change in the capital/labor ratio greater? Which will have a higher elasticity of substitution?
3 Determine the elasticity of substitution in the case of the isoquant in panel A of Figure 6.1.
4 Suppose a firm has the production technology shown below for Goods 1 and 2.
   (a) Does Good 1 indicate economies of scale? Why?
   (b) Does Good 2 indicate economies of scale? Why?
   (c) Do the two goods indicate economies of scale? Why?

Good 1		*Good 2*		*Both*		
Q_1	*Cost*	Q_2	*Cost*	Q_1	Q_2	*Cost*
10	50	10	60	10	10	100
20	100	20	100	20	20	180
30	150	30	130	30	30	250

5 If any firm's price of labor and capital each double, what will happen to the expansion path (i.e., locus of tangencies between the isoquants and isocost curves)? What will happen to the firm's average cost curve?
6 Figure 6.5, panel A, illustrates technical inefficiency for firms with a one-input production function. It was explained that inefficiency could be measured by output distance or, alternatively, input reduction distance. Sketch and explain the comparable measures for the two-input production function.
7 In Escarce's account of diffusion, do improved "channels of information" matter regardless of the information content? Does all information increase the adoption rate? If not, what information does?
8 Calculate the average costs at points *C*, *F*, and *G* in Figure 6.2. Do they imply increasing or decreasing returns to scale? Why?

Notes

1 The new ratio is 99 physicians to 105 nurses, or 0.94.
2 Work by Olesen and Petersen (2002) promises to provide ways to incorporate large numbers of hospital outputs into fewer output measures.
3 Hornbrook and Monheit (1985) study the importance of case mix. From data for 380 hospitals, they found that larger-scale hospitals in their sample tended to admit case mixes with relatively shorter lengths of stay.
4 See Fare and Lovell (1978) for economic applications of the DEA approach.
5 Psychiatric hospitals (Mark, 1996), nursing homes (Kooreman, 1994a; Vitaliano and Toren, 1994); group homes (van Lear and Fowler, 1997), physician clinics (Defelice and Bradford, 1997; Gaynor and Pauly, 1990), and physicians working in hospitals (Chilingerian, 1995) also have been studied.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 7

Demand for Health Capital

In this chapter

- The Demand for Health
- Labor–Leisure Trade-Offs
- The Investment/Consumption Aspects of Health
- Investment over Time
- The Demand for Health Capital
- Changes in Equilibrium: Age, Wage, and Education
- Empirical Analyses Using Grossman's Model
- Obesity—The Deterioration of Health Capital
- Conclusions

Chapter 5 considered the production of health in the aggregate by looking at the impacts of various factors on health for society as a whole. Here, we show how individuals allocate their resources to produce health. Economists' understanding of this decision has been deepened by the important work of Michael Grossman and his followers. The model has enabled us to understand thoroughly the roles of age, education, health status, and income in the production of health through the demand for health capital. It also provides a useful format for examining the causes and impacts of obesity (being overweight), an important and current family health topic.

The Demand for Health

The Consumer as Health Producer

Grossman (1972a, 1972b) extended the theory of human capital to explain the demand for health and health care. According to human capital theory, individuals invest in themselves through education, training, and health to increase their earnings. Health investment leads to a myriad of decisions regarding work, heath (and non-health) care, exercise, and the use of consumption goods and bads (such as unhealthy food, cigarettes, alcohol, or addictive drugs). In particular, according to this theory:

1 It is not medical care as such that consumers want, but rather health. In seeking health; they demand medical care inputs to produce it.
2 Consumers do not merely purchase health passively from the market. Instead, they produce health, combining time devoted to health-improving efforts including diet and exercise with purchased medical inputs.
3 Health lasts for more than a day or a month or a year. It does not depreciate instantly, and it can be analyzed like a capital good.
4 Perhaps most importantly, health can be treated both as a consumption good and an investment good. People desire health as a consumption good because it makes them feel and look better. As an investment good, they desire health because it increases the number of healthy days available to work and to earn income.

Figure 7.1 provides a simple diagram that explains the concept of health capital. Just as cars or refrigerators, as capital goods (or "stock of capital") provide streams of services over time, one can conceive of a person's stock of health capital that provides the ultimate output of "healthy days." One might measure the stream of health output(s) as a single dimension of healthy days, or in several dimensions of physical health, mental health, and limited activity; for example, one can no longer play singles in tennis, but must play (less strenuous) doubles instead.

Consumers apply sets of health inputs, which might include not only market inputs of health care, but also diet, exercise, and time, to their physical makeup, thus making investments in health capital. These investments maintain or improve the consumers' stocks of health, which in turn provide them with healthy days. Over time, the health stock may grow, remain constant, or decline (again, like a car or a refrigerator), either slowly with age, or more precipitously with illness or injury. As noted in Box 7.1, there may be many technologies available to produce health capital, using various amounts of time or market goods.

From Figure 7.1 and the accompanying discussion, we see how the end goal of "healthy days" guides consumer decisions as to how much time and money to invest in health stock. We will see that the prices of health care, the people's wage rates, and their productivities in

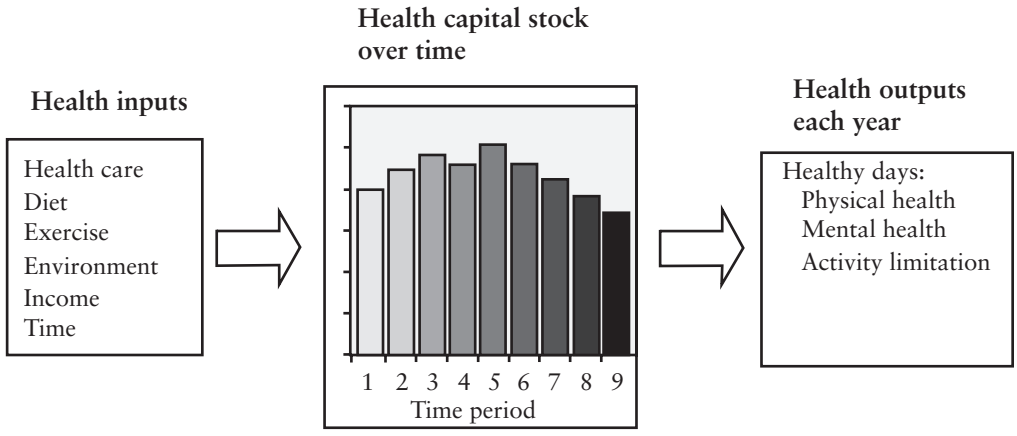


Figure 7.1 Investing in Health Capital

producing health will dictate how resources are to be allocated among health capital, and other goods and services that people buy.

Consider a consumer, Ed Kramer, who buys market inputs (e.g., medical care, food, clothing), and combines them with his own time to produce a stock of health capital that produces services that increase his utility. Ed uses market inputs and personal time both to invest in his stock of health and to produce other things that he likes.

These other items include virtually all other things that Ed does. They include time spent watching television, reading, playing with and teaching his children, preparing meals, baking bread, or watching the sun set, a composite of other things people do with leisure time. We shall call this composite home good B.

Time Spent Producing Health

An increment to capital stock, such as health, is called an investment. During each period, Ed produces an investment in health, I. Health investment I is produced by time spent improving health, T_H, and market health inputs (providers' services, drugs), M. Likewise, home good B is produced with time, T_B, and market-purchased goods, X.

BOX 7.1

Exercise Technology—FitBits or Smartphones?

Exercise provides a valuable input into the production of health, and exercisers have improved access to a myriad of technological aids to monitor and improve their performance. Many of these aids cost $100 or more. Are they worth it?

A research team at The University of Pennsylvania wondered the same thing. They developed a study to test 10 of the top-selling smartphone apps and devices in the United States by having 14 participants walk on a treadmill for 500 and 1,500 steps,

each twice (for a total of 56 trials), and then recording their step counts. Study participants, all healthy Penn adults, wore the following devices during the treadmill trials:

- Waistband: one pedometer and two accelerometers.
- Wrists: three wearable devices.
- Pants pockets: two smartphones, one running three apps and the other running one.

The data from the smartphones were only slightly different than the observed step counts (with a range of –6.7 to +6.2 percent relative difference in mean step count), but the data from the wearable devices differed more (with a range of –22.7 to –1.5 percent). In short, there was no systematic advantage to using dedicated wearable devices.

The takeaway is that while smartphones might be a little bulky for vigorous exercise, "increased physical activity facilitated by these devices could lead to clinical benefits not realized by low adoption of pedometers." The authors felt that their findings may help reinforce individuals' trust in using smartphone applications and wearable devices to track health behaviors, which could have important implications for strategies to improve population health.

Source: Case, Meredith A., et al., "Accuracy of Smartphone Applications and Wearable Devices for Tracking Physical Activity Data," *Journal of the American Medical Association* 313(6) (2015): 625–626. doi:10.1001/jama.2014.17841, accessed January 28, 2016.

If, for example, we considered good B to be baking bread, the market goods might include flour, yeast, kitchen appliances, and gas, water, and/or electricity. Thus, Ed uses money to buy health care inputs, M, or home good inputs, X. He uses leisure time either for health care (T_H) or for producing the home good (T_B). Using functional notation:

$$I = I\,(M, T_H) \tag{7.1}$$

$$B = B\,(X, T_B) \tag{7.2}$$

These functions indicate that increased amounts of M and T_H lead to increased investment I, and that increased amounts of X and T_B lead to increased home good B.

In this model, Ed's ultimate resource is his own time. Treat each period of analysis as being a year, and assume that Ed has 365 days available in the year. To buy market goods such as medical care, M, or other goods, X, he must trade some of this time for income; that is, he must work at a job. Call his time devoted to work T_W.

Because our focus is on the health aspects of living, we realize that some of his time during each year might involve ill health, or T_L. Thus, we account for his total time in the following manner:

$$\begin{aligned}\text{Total time} = T = 365 \text{ days} = T_H\,(\text{improving health}) + T_B\,(\text{producing home goods}) \\ + T_L\,(\text{lost to illness}) + T_W\,(\text{working})\end{aligned} \tag{7.3}$$

Recall that his leisure time is spent either improving his health or producing home goods.

Labor–Leisure Trade-Offs

The labor–leisure trade-off illustrates the potential uses of Ed's time. Our variation on this analysis also helps illustrate the investment aspects of health demand.

Trading Leisure for Wages

In Figure 7.2, the *x*-axis represents Ed's work and leisure time. Suppose that he considers his time spent creating health investment to be "health-improvement time" and that he calls T_B his leisure. In reality, he may do some health-improving activities at work, may obtain some enjoyment or satisfaction from healthful time, and so on, but assume here that these categories are exclusive. Assume further that the number of days lost to ill health and the number of days spent on health-enhancing activity are fixed (we relax this assumption later). Variables T_L and T_H refer to time lost and time spent on healthy activities, respectively. The maximum amount of time that he has available to use either for work, T_W, or leisure, T_B, is thus $365 - T_H - T_L$, so:

$$\text{Time Available for Work or Leisure} = 365 \text{ days} - T_H - T_L = T_B + T_W \quad (7.4)$$

Leisure time, T_B, is measured toward the right while time spent at work, T_W, is measured toward the left. Figure 7.2 shows that if Ed chooses leisure time, OA, then he has simultaneously chosen the amount of time at work indicated by AS.

Recall that Ed's total amount of time available for either work or leisure is given by point *S*. If he chose point *S* for the period, he would be choosing to spend all this available time in leisure; that is, in the pursuit of the pleasures of life (albeit without the wage income to produce them). The *y*-axis represents income, obtained through work. This income will then purchase either market health goods or other market goods. Thus, if he chooses point *S*, he will not be able to purchase market goods because he has no wage income.[1]

If, beginning at *S*, Ed gives up one day of leisure by spending that day at work, to point *N*, he will generate income equal to $0Y_1$, which represents his daily wage. In economic terms,

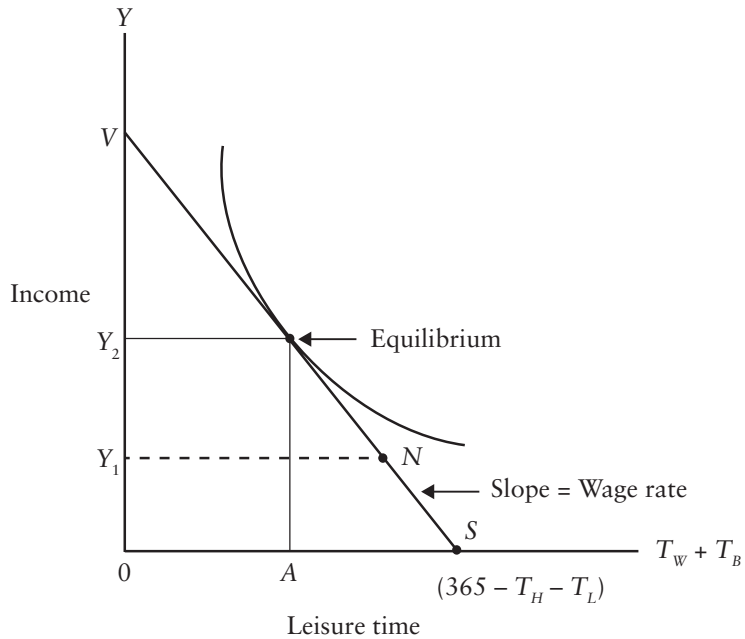


Figure 7.2 Labor–Leisure Trade-Off

this quantity represents income divided by days worked—that is, the daily wage. The slope of the line *VS* depicting the labor–leisure trade-off reflects the wage rate (if Ed pays Social Security and/or taxes on his wage, then the slope reflects the after-tax wage rate).

Preferences between Leisure and Income

Consumers have preferences regarding income and leisure, just as they had among other goods in Chapter 2. As before, Ed would like more income and more leisure so the indifference curve map is shaped normally. In Figure 7.2, in equilibrium, Ed's trade-off of leisure and income is the same as the market trade-off, which is the wage rate. Here, he takes amount $0A$ of leisure and trades amount AS of leisure for income, $0Y_2$.

In Figure 7.3, Ed has made a different choice with respect to time spent investing in health status. To illustrate, suppose that time spent on health-producing activities, T_H, is increased to T'_H. Correspondingly, suppose that the number of days lost to ill health has been reduced to T'_L. What effect will this change in time have on the horizontal intercept, which is the total time remaining for work or leisure? On the one hand, the time he spends producing health reduces his time available for other activities. Time spent on health investment increases health stock and, in turn, reduces time lost to illness.

If the net effect of $T'_H + T'_L$ is a gain in available time, then this illustrates the pure investment aspect of health demand. The health investments "pay off" in terms that both add to potential leisure and also increase the potential income, shifting the income–leisure line outward from *VS* to *RQ*. The expenditure of time (and medical care, too) for health-producing activities may later improve Ed's available hours (because he is sick less) of productive activity.

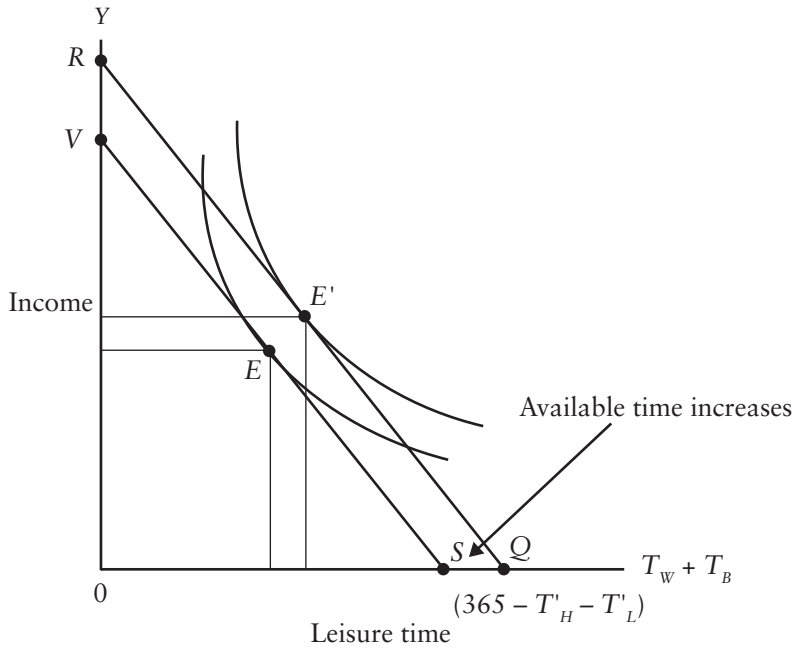


Figure 7.3 Increased Amount of Healthy Time Due to Investment

As a result of his investment, Ed can increase his utility, moving from point E to point E'. Not only does investment in health lead to his feeling better, but it also leads to more future income and may lead to more leisure, as well.

The improvement in health status also might increase Ed's productivity at work, perhaps resulting in a higher wage and a steeper income–leisure line (why is it steeper?). In any case, the analysis shows that Ed might wish to engage in activities to improve his health, even if the only value of health is its effect on his ability to earn future income.

The Investment/Consumption Aspects of Health

The Grossman model describes how consumers simultaneously make choices over many periods or years. It can also be instructive to represent a whole life span as a single period. This can show the dual nature of health as both an investment good and a consumption good.

Production of Healthy Days

For simplicity, we will view health as a productive good that produces a single output, healthy days, a production function relationship illustrated in Figure 7.4. The horizontal axis measures health stock in a given period. A larger stock of health leads to a larger number of healthy days, up to a natural maximum of 365 days. The bowed shape of the curve illustrates the law of diminishing marginal returns. That is, additional resources have decreasing marginal impacts on the output—all of Ed's health improvement attempts serve to increase his healthy days, but he still gets sick every once in a while. Note also the concept of a health stock minimum shown as H_{min}, where health stock becomes arbitrarily small. At this point, production of healthy days drops to zero, indicating death.

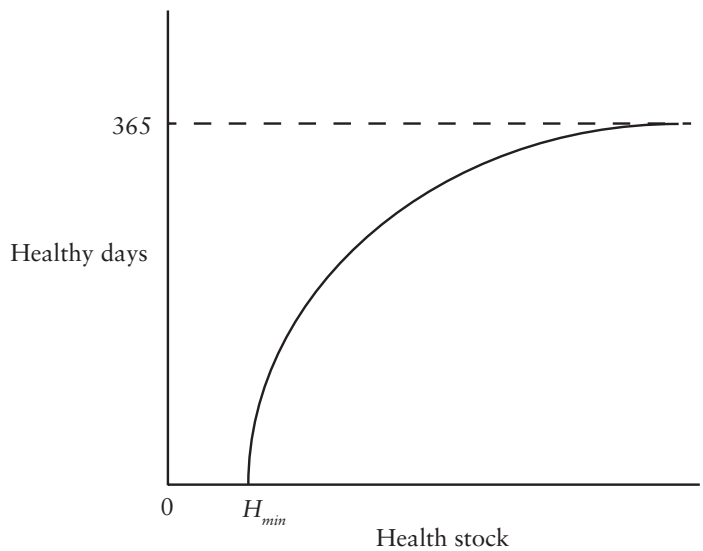


Figure 7.4 Relationship of Healthy Days to Health Stock

Production of Health and Home Goods

Consider the possibilities for producing health, H, and home good, B, given the total amount of time available. Figure 7.5 shows the production possibilities trade-off. The curve differs from the usual production possibilities curve in several respects. First, from point A to point C, health improvements increase the amounts of the home good, B, and health, H, attainable. It is necessary to increase health beyond H_{min} in order to obtain income and leisure time from which to produce B.

Moving along the production possibilities curve, Ed shifts his uses of available time and distributes his purchases of market goods. The move from E to C indicates that he has made more time available for health and that this move has reaped the side benefit (increased leisure time) of increasing the availability of market goods and time used to increase production of the home good, baking bread.

Suppose that Ed desires health solely for its effect on the ability to produce income and the leisure time to produce the home good bread, B. This would imply that his indifference curves between H and B are vertical lines. (Ed places no intrinsic value on H, so he would not trade any B to get additional health.) In such a case, he would maximize his utility by producing as much B as possible. The utility-maximizing choice would be at point C, a point of tangency between indifference curve U_1 and the production possibilities curve. He produces amount B_0 of the home good, and H_0 of health.

Now assume instead that Ed achieves utility not only from producing B, but also directly from health itself (he likes feeling better). In this case, his (dashed) indifference curve, U_2, has the more familiar curved shape in Figure 7.5, sloping downward from left to right. It is more realistic to say that he values health both as a consumption good, as is shown in Figure 7.5, and as an investment in productive capacity. The consumption aspect suggests that he enjoys feeling healthy; the investment aspect, that feeling healthy makes him more productive, thus allows him to earn more. In general, by including Ed's "feeling healthy" in this consumption feature of the model, he will choose a higher health stock than under the pure investment model. In Figure 7.5, health stock, H_1, exceeds H_0. The cost of this increase in H involves foregoing some of the home good B, such that B_1 is less than B_0.[2]

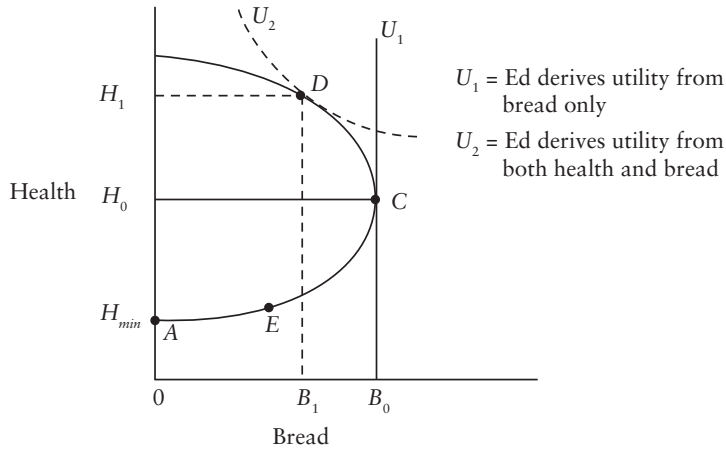


Figure 7.5 Allocation of Production between Health and Bread

Investment over Time

The Cost of Capital

People make choices for the many periods over their life cycles, rather than just for one representative period. As a beginning point for each analysis, we feature the pure investment version of the model (point *C* in Figure 7.5). We then discuss the analytical changes when consumers, in addition, value health intrinsically (point *D* in Figure 7.5). We demand health capital because it helps us earn more and feel better. What does it cost? By analogy, a health clinic purchases thousands of dollars of X-ray equipment. The return to the X-ray equipment is in the future earnings that ownership of the equipment can provide.

Suppose that an X-ray machine costs $200,000, and that its price does not change over time. Suppose also that the annual income attributable to the use of the X-ray machine is $40,000. Is purchasing the machine a good investment? Consider the alternative: Instead of purchasing the X-ray machine, the clinic could have put the $200,000 in a savings account, at 5 percent interest, yielding the following:

$$200{,}000 \times 1.05 = 210{,}000 \text{ at the end of Year 1}$$

$$200{,}000 \times 1.05^2 = 210{,}000 \times 1.05 = 220{,}500 \text{ at the end of Year 2}$$

$$200{,}000 \times 1.05^3 = 220{,}500 \times 1.05 = 231{,}525 \text{ at the end of Year 3}$$

$$200{,}000 \times 1.05^4 = 231{,}525 \times 1.05 = 243{,}101 \text{ at the end of Year 4}$$

$$200{,}000 \times 1.05^5 = 243{,}101 \times 1.05 = 255{,}256 \text{ at the end of Year 5}$$

For the investment in an X-ray machine to be desirable by these criteria, it should provide at least $55,256 in incremental revenue over the five years.

The problem is more complicated, however, because capital goods depreciate over time. Most students will agree that a five-year-old personal computer is worth almost nothing. Even though it may do everything it ever did, new programs may not work on it, new equipment may not hook up to it properly, and it may be very slow compared to new machines. In economic terms, the machine has *depreciated*, and if parts like disk drives wear out, it may depreciate physically as well. If the computer originally cost $2,000, and the students have not budgeted $2,000 for replacement, they may find themselves without working computers! In economic terms, depreciation is the budgeting for replacement.

Similarly, suppose that the clinic knows that the X-ray machine will wear out (or depreciate), so that it will be worth only half its original value in five years. The clinic must earn enough not only to cover the opportunity cost from the bank, but also to maintain the value of the machine. For an investment in an X-ray machine to be worthwhile, then, it must not only earn the competitive 5 percent return each year, but it also must provide enough return to cover depreciation. This suggests that the cost of holding this capital good for any one year, as well as over time, will equal the opportunity cost of the capital (interest foregone) plus the depreciation (deterioration of value). In this example, the depreciation cost is $100,000, or half of the $200,000 original cost.

The Demand for Health Capital

Conventional economic analysis provides a powerful conceptual apparatus by which to analyze the demand for a capital good. The cost of capital, in terms of foregone resources

(for health capital, both time and money), is a supply concept. The other needed tool is the concept of the marginal efficiency of investment (*MEI*), a demand concept that relates the return to investment to the amount of resources invested.

Marginal Efficiency of Investment and Rate of Return

The *MEI* can be described in terms of the X-ray machine example. A busy clinic may wish to own more than one X-ray machine. But how many? The clinic management may logically consider them in sequence. The first machine purchased (if they bought only one) would yield a return as we have discussed. Suppose that return each year is $40,000.

We also can calculate the annual *rate of return*, which would be $40,000 ÷ $200,000, or 20 percent per year. The management would buy this machine if the incremental revenue brought in covered its opportunity cost of capital and the depreciation. In terms of rates, management would choose to own the first X-ray machine as long as the rate of return, 20 percent, exceeded the interest rate (the opportunity cost of capital) plus the depreciation rate.

If management considered owning two machines, it would discover that the rate of return on the second X-ray machine would probably be less than the first. This is best understood by recognizing that a clinic buying only one X-ray machine would assign it to the highest-priority uses, those with the highest rate of return. If the clinic were to add a second X-ray machine, then logically it could be assigned only to lesser-priority uses (and might be idle on occasion). Thus, the second machine would have a lower rate of return than the first. The clinic would then purchase the second machine only if its rate of return were still higher than interest plus depreciation.

The Decreasing MEI

Other machines probably could be added at successively lower rates of return. In Figure 7.6, the marginal efficiency of investment curve, *MEI*, describes the pattern of rates of return, declining as the amount of investment (measured on the horizontal axis) increases. The cost

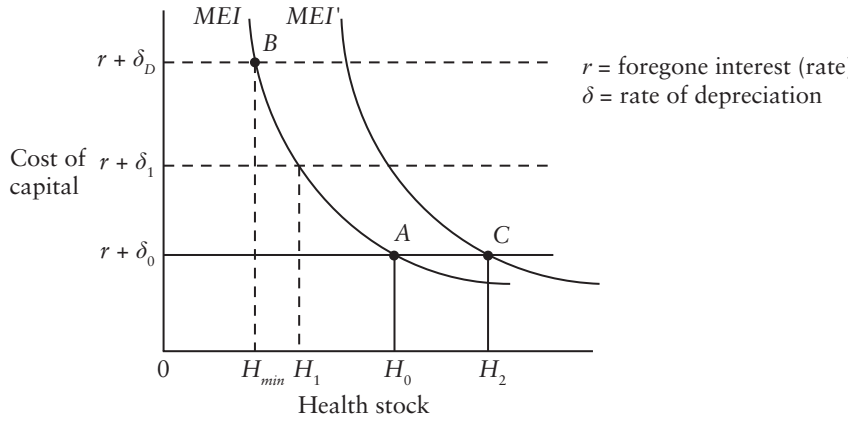


Figure 7.6 Optimal Health Stock

of capital (that is, the interest rate, r, plus the depreciation rate, δ_0) is shown as the horizontal line labeled $r + \delta_0$. The optimum amount of capital demanded is thus H_0, which represents the amount of capital at which the marginal efficiency of investment just equals the cost of capital. This equilibrium occurs at point A.

Like the marginal efficiency of investment curve in this example, the *MEI* curve for investments in health also would be downward sloping. This occurs because the production function for healthy days (see Figure 7.4) exhibits diminishing marginal returns. The cost of capital for health would similarly reflect the interest rate plus the rate of depreciation in health. Understand that a person's health, like any capital good, also will depreciate over time. As we age, certain joints may wear out, certain organs may function less well than before, or we may become more forgetful. Thus, the optimal demand for health is likewise given at the intersection of the *MEI* curve and the cost of capital curve ($r + \delta_0$).

Changes in Equilibrium: Age, Wage, and Education

Return to Ed Kramer and to Figure 7.1. Based on the analyses thus far, Ed has chosen an equilibrium level of health stock, by deciding how much to work, how much time to spend on health, what kind of diet, and how much exercise to do. He allocates his resources such that every year he maintains a constant level of health stock, and this provides him with an equilibrium level of healthy days per year. How does his investment in health change in response to changes in age, wage, and education? The model depicted in Figure 7.6 helps us to investigate several important model implications. Consider age first.

Age

How does Ed's optimal stock of health vary over a lifetime? In this model the age of death is determined as part of the model; death itself is *endogenous*, meaning that it doesn't just happen! (For completeness, an outcome determined outside of the model is termed *exogenous*.) Here, Ed chooses his optimal life span, a life span that is not infinite. By this model, all of us at some time will optimally allow our health stock to dissipate to h_{min}. This feature depends in a critical way on how the depreciation rate (a cost factor) varies with age, as well as how long the person expects to live (and enjoy the benefits of good health).

Looking first at costs, Ed's health stock may depreciate faster during some periods of life and more slowly during others. Eventually, as he ages, the depreciation rate will likely increase. In other words, the health of older people will likely deteriorate faster than the health of younger people.

Consider then the effect of aging on Ed's optimal health stock. Return to Figure 7.6. We assume that the wage and other features determining the *MEI* are not substantially altered by this aging. However, by hypothesis, the depreciation rate, δ, increases with age from δ_0 to δ_1 and ultimately to δ_D. These assumptions imply that the optimal health stock decreases with age.

This situation is shown in Figure 7.6 by the fact that the optimal health at the younger age, H_0, is greater than H_1, the optimal stock at the older age. Higher depreciation rates increase the cost of holding health capital stock. We adjust to this by holding more health capital in periods when health is less costly. In old age, health depreciation rates are extremely high, δ_D, and optimal health stock falls to H_{min} at point B.

This conclusion is consistent with the observation that elderly people purchase a greater amount of medical care, even as their health deteriorates. Grossman explains the phenomenon:

> Gross investment's life cycle profile would not, in general, simply mirror that of health capital. This follows because a rise in the rate of depreciation not only reduces the amount of health capital demanded by consumers but also reduces the amount of capital supplied to them by a given amount of gross investment.
>
> (p. 238)

Turning to returns from investment, consider an analogy to a consumer who has two cars, the same models, built by the same maker. One is two years old, with 30,000 miles of wear; the other is 12 years old, with 180,000 miles of wear. Each car has had the "bumps" and "bruises" that accompany driving in a major city. Suppose that the two-year-old car is damaged in the parking lot and will cost $3,000 to repair. The decision is obvious. However, would the owner incur the same level of costs (more pointedly, would the car insurer reimburse these costs) to repair the 12-year-old car, which is closer to the end of its useful life?

The analogy to individual health is immediate. A younger person may choose complicated surgery to replace knees or hips, to maintain earning capability or quality of life. An older person may choose not to do so. In Figure 7.6, this suggests that as the expected length of life decreases, the *MEI* curve shifts to the left, because the returns from an investment will last for a shorter period of time. This will reinforce the decrease in investment that occurs due to increased depreciation.

Other features of the model suggest that people will increase their gross investment (their expenditures) in health as they age. It follows, in turn, that the elderly would demand more medical care than the young, as we frequently note to be the case.

Thus, the pure investment model generates the prediction that optimal health will decline as the person ages. Will this prediction change when we assume more realistically that an individual also will value health for consumption reasons (good health makes one feel better)? The issue relies on whether older persons get more or less direct utility from the enjoyment of healthy days. If people increase their valuation of healthy days as they age, this may offset the predicted health stock decline.

Wage Rate

Figure 7.6 also illustrates the effect of a change in the wage rate on Ed's optimal level of investment. Increased wage rates increase the returns obtained from healthy days (8 hours' work will bring in $160 rather than $120 if the wage rate increases from $15 to $20 per hour). Thus, higher wages imply a higher *MEI* curve, or *MEI′*.

Assume now that the original *MEI* curve describes the lower-wage case and yields optimal health stock, H_0. The *MEI′* curve, above *MEI*, shows the marginal efficiency of investment for someone with higher wages. At new equilibrium point *C*, the higher wage will imply a higher optimal level of health stock, H_2, in this pure investment model. The rewards of being healthy are greater for higher-wage workers, so increased wages will generally tend to increase the optimal capital stock.[3]

The model illustrates one more implication of the wage factor. Consider that when Ed retires, his wage effectively drops to zero because improved health does not help him earn more. The pure investment version implies that he would change his optimal health stock to

H_{min} upon retirement. Once he retires, he would make no further investment in health, but instead would allow health to depreciate until death.

How would we amend this analysis by considering the consumption aspects of health—that good health makes people feel good? First, Ed would presumably continue to obtain utility directly from healthy days. Thus, optimal health stock would not necessarily drop to H_{min} directly upon Ed's retirement, but it would do so only when depreciation rates became sufficiently severe.

Second, if retirees and those who are still working obtain utility directly from healthy days, then the only significant change upon retirement would involve the end of the pure investment motive. Therefore, even when we include the consumption aspects of health, we would expect people to reduce their health stock upon retirement.

Education

Education is especially interesting to those who study health demand. Those with higher education often have better health, and most economists believe that education may improve the efficiency with which people can produce investments to health and the home good. Examples of improved efficiency may include improved ability to follow instruction regarding medicines or better knowledge of harmful effects of obesity, or of potentially harmful activities such as smoking, drinking, or addictive drugs, to name just a few.

Figure 7.6 illustrates the effect of education. Here, the *MEI* curve illustrates the marginal efficiency of investment for the consumer with a low level of education (measured, for example, by years of schooling), while the *MEI'* curve illustrates the same person with a higher level of education. This model indicates that because education raises the marginal product of the direct inputs, it reduces the quantity of these inputs required to produce a given amount of gross investment.

It follows that given investments in health can be generated at less cost for educated people, and thus they experience higher rates of return to a given stock of health. The result, as shown, is that the more highly educated people will choose higher optimal health stocks, H_2; the less highly educated will choose H_0.

This explains the widely observed correlation between health status and education. Educated people tend to be significantly healthier. However, this explains only the correlation of health status and education from the supply side in that it considers only the increased efficiency with which we produce health. One also might wish to explain education from the demand side.

Educated people most likely recognize the benefits of improved health. They may enjoy preparing and eating nutritious food or doing physical exercise. They may recognize the dangers of smoking and the long-term problems of overexposure to the sun. They may enjoy feeling and looking good. As such, all else equal, they would have a greater taste for health relative to other goods.

The demand for health due to education is difficult to separate from the supply effect of education, which implies more productivity in producing health. Clearly, however, both exist and both are important.

Empirical Analyses Using Grossman's Model

Hundreds of articles have looked to Grossman's model for guidance in examining health production, and Box 7.2 discusses briefly the important insights related to addictive behavior.

The resulting literature examines the separable impacts of age and education. It also looks at the various time components of health investment within families, and the impacts of health status on the demand for health and health care. We sample several studies here, and look at particular recent examples relating to obesity in the next section.

Sickles and Yazbeck (1998) developed a structural model of the demand for leisure and the demand for consumption for elderly males. Measuring health production is a difficult problem, so the authors use the Quality of Well-Being (QWB) scale, developed by Kaplan and Anderson (1988), based on mobility, physical activity, social activity, and physical symptoms and problems. They find that both health care and leisure consumption tend to improve health. A 1 percent increase in health-related consumption increases health by 0.03 to 0.05 percent. A 1 percent increase in leisure increases health by from 0.25 to 0.65 percent.

BOX 7.2

Rational Addiction

Economists Gary Becker and Kevin Murphy (1988) describe conditions under which rational people choose to consume addictive goods. Their work challenges our common moral concepts and attitudes toward addiction. It also offers similarities and contrasts with Grossman's model of the demand for health.

Becker and Murphy argue that addictive behavior (as most current or former smokers would agree) must provide some pleasure or people would not pursue it. For a good to be addictive, in their model, past consumption (of the good) must increase the marginal utility of current consumption. Past consumption contributes to "consumption capital" of the good in question. Current smoking may entail a learning experience in the sense that future consumption of an additional cigarette becomes more enjoyable. So, also, do drinking coffee or listening to Mozart. Mozart symphonies may become more enjoyable in the future as this learning experience continues. Coffee drinkers learn to enjoy the beverage but discover that stopping usage can lead to symptoms of withdrawal.

Many addictions, however, are harmful. Harmful, in the Becker and Murphy context, means that the capital good—the consumption capital—has harmful effects similar to a reduction in health status in the Grossman model. Smoking cigarettes may reduce healthy days and may reduce income. Such harmful effects are part of what Becker and Murphy call the "full price" of the addictive good.

A potential cigarette smoker or drug user might choose to start, even knowing that it is addictive, if smoking or using tends to increase future enjoyment from smoking more than it increases future harmful effects. This explains why a rational person may choose to ingest a harmful substance, knowing fully its harmful effects. (See Chapter 23 for a more detailed discussion of rational addiction.)

Gerdtham and Johannesson (1999) estimate health demand with a Swedish sample of over 5,000 adults using a categorical measure of overall health status to measured health capital. They find that the health demand increases with income and education and decreases with age, urbanization, being overweight, and being single.

Another way to test the model involves the demand for preventive care. Kenkel (1994) estimates the determinants of women's purchases of medical screening services, designed for

the early detection of breast and cervical cancer. Annual use of these screening tests decreases with age, a result consistent with women's rationally reducing care as the payoff period shortens over the life cycle and as the depreciation rate rises. Furthermore, increased schooling tends to increase the use of the screening services, implying more efficiency in producing health.

Mullahy and Sindelar (1993) examine the relationships among alcoholism, income, and working. Poor health may reduce income either by reducing productivity, which results in lower wages, or by reducing labor market participation (whether and/or how much one is working). Alcoholism is an illness that reduces people's health capital. Mullahy and Sindelar find the labor market participation effects are more powerful than the wage (productivity) effects in reducing earnings, and hence reducing the return to health. By these criteria, successful alcoholism treatment would appear to have significantly positive economic returns.

In a retrospective essay on research achievements and directions in the 30 years after Grossman's pioneering analysis, Leibowitz (2004) finds that increases in the parents' valuations of time will also affect the relative costs of alternative inputs to children's health. As a result, mothers who work outside the home may employ substitutes for their own time that are less effective in producing child health. For example, working mothers can substitute prepared foods for their own time in producing meals for their children. However, these prepared foods are often high-calorie and high-fat, perhaps leading to less nutritious diets for their children (Leibowitz, 2003; Anderson et al., 2003).

In yet another application, Borisova and Goodman (2004) examine the importance of time in compliance of drug abusers with methadone treatment as a substitute for heroin. Because most drug abuse treatment comes at only nominal or zero cost, the time spent traveling to, and receiving, the daily treatment becomes critical. The authors find, as expected, that increased travel and treatment time costs significantly reduce treatment compliance, and decrease the improvement in health capital related to kicking the addiction.

Bhattacharya and Lakdawalla (2006) examine the value to the U.S. labor force of improvements in survival and health over the years 1970 to 1999. They find that survival gains and reductions in the number of work-days missed due to poor health have added about 8 percent to the remaining labor force value of black males, and about the same to the value of 60-year-old white males. They note that these effects are almost as large as a full year of schooling. Gains for younger white males appear to have been approximately 5 percent, and gains for women are around 2 percent. Overall, they estimate that health improvements have added $1.5 trillion to the value of labor market human capital over this period. Even a 5 percent return on this human capital would lead to an increase of $75 billion per year in GDP, a substantial amount!

All of these examples indicate that Grossman's model has yielded considerable insight into the determinants of health and into the allocation of resources (both time and money) into health-creating activities. It has also crossed fields of economics, including labor, development, and growth economics, and has provided fruitful results in cross-disciplinary demographic and sociological research as well.

Obesity—The Deterioration of Health Capital

Obesity (excess weight) provides many insights in a model of health capital. Aside from aesthetic issues regarding appearance, obesity is a leading risk factor for heart disease, hypertension (high blood pressure), certain cancers, and type-2 diabetes.

According to reports from the Center for Disease Control in 2016, over almost 35 percent of U.S. adults (almost 79 million people) and 17 percent of U.S. children were obese in 2011–2012. From 1980 through 2012, obesity rates for adults and children have doubled and rates for adolescents have quadrupled. During the past several decades, obesity rates for all groups in society regardless of age, sex, race, ethnicity, socioeconomic status, education level, or geographic region have increased markedly.

Health analysts usually measure obesity in terms of Body Mass Index, or BMI, with the formula BMI = Weight in kilograms / (Height in meters)^2. It is not a complete measure. For example, BMI does not distinguish between fat and muscle, and entire rosters of U.S. professional football teams would be termed as overweight or obese by the BMI. Nonetheless it is easy to use, and it adjusts for the fact that tall people are generally heavier. If Ed Kramer weighs 90 kg (almost 200 pounds) and is 1.75 meters (about 5 feet 9 inches) tall, he has a BMI of 29.4. Table 7.1 provides the generally accepted measures of weight relative to body size measured by height.

As noted in Table 7.1, a BMI rating of over 25 is classified as overweight. By this criterion, Ed is overweight, and if he weighed 2 kilograms more, with a BMI slightly over 30, he would be considered (Class I) obese.

Table 7.2 indicates that 44 states in 2014 in the United States had obesity prevalence (BMI greater than 30) equal to or greater than 25 percent of their adult populations, 19 had prevalence between 30 and 35 percent, and three states (Arkansas, Mississippi, and West Virginia) had prevalence greater than 35 percent.

Table 7.1 Weight Status Classified by Body Mass Index

Category	*BMI range*
Severely underweight	Less than 16
Underweight	16 to 18.5
Normal	18.5 to 25
Overweight	25 to 30
Obese Class I	30 to 35
Obese Class II	35 to 40
Obese Class III	40 and above

Source: World Health Organization, www.who.int/bmi/index.jsp?introPage=intro_3.html, accessed November 2016.

Table 7.2 2014 U.S. State Obesity Rates

Alabama	33.5	Illinois	29.3	Montana	26.4	Rhode Island	27.0
Alaska	29.7	Indiana	32.7	Nebraska	30.2	South Carolina	32.1
Arizona	28.9	Iowa	30.9	Nevada	27.7	South Dakota	29.8

Table 7.2 *continued*

Arkansas	35.9	Kansas	31.3	New Hampshire	27.4	Tennessee	31.2
California	24.7	Kentucky	31.6	New Jersey	26.9	Texas	31.9
Colorado	21.3	Louisiana	34.9	New Mexico	28.4	Utah	25.7
Connecticut	26.3	Maine	28.2	New York	27.0	Vermont	24.8
Delaware	30.7	Maryland	29.6	North Carolina	29.7	Virginia	28.5
Washington, DC	21.7	Massachusetts	23.3	North Dakota	32.2	Washington	27.3
Florida	26.2	Michigan	30.7	Ohio	32.6	West Virginia	35.7
Georgia	30.5	Minnesota	27.6	Oklahoma	33.0	Wisconsin	31.2
Hawaii	22.1	Mississippi	35.5	Oregon	27.9	Wyoming	29.5
Idaho	28.9	Missouri	30.2	Pennsylvania	30.2		

Source: www.cdc.gov/obesity/data/table-adults.html, accessed January 27, 2016.

Table 7.3 Percent BMI > 30, 2014

a. Selected Countries

Country	*Both sexes*	*Female*	*Male*
Argentina	26.3	28.9	23.6
Australia	28.6	28.8	28.4
Brazil	20.0	22.7	15.1
Canada	25.9	27.2	24.6
Egypt	26.2	34.8	17.7
France	22.0	22.3	21.8
Germany	18.5	17.2	19.9
Ireland	25.6	25.3	25.9
Israel	25.3	27.0	23.5
Italy	21.0	21.6	20.4
Japan	3.3	3.2	3.4
Mexico	25.9	30.7	20.7
Nigeria	8.9	13.4	4.6
Poland	23.1	25.1	21.0
Russian Federation	22.2	26.2	17.6
United Kingdom	25.5	26.8	24.1
United States	31.2	32.5	29.8

Table 7.3 *continued*

b. Highest Obesity Values

Country	*Both sexes*	*Female*	*Male*
Cook Islands	48.2	53.1	43.5
Palau	46.1	51.0	41.4
Nauru	45.4	51.0	39.9
Samoa	41.8	50.0	34.0
Marshall Islands	41.7	47.8	35.6
Tonga	41.6	49.0	34.3
Niue	41.0	47.3	35.0
Kiribati	39.2	47.2	31.5
Tuvalu	38.4	44.6	32.4
Qatar	38.1	46.5	35.4
Kuwait	36.8	43.2	32.5
Micronesia	35.6	42.4	29.2
Fiji	35.0	40.9	29.4

Source: http://apps.who.int/gho/data/view.main.2450A, accessed January 27, 2016.

Obesity is not limited to the United States. Table 7.3a (the age-adjusted obesity rate) shows that many advanced countries have well over one-quarter of their populations with BMI greater than 30. Typically, although not always, female rates exceed male rates. The highest obesity rates (as noted in Table 7.3b) occur in several Pacific Island nations, as well as some Middle Eastern states.

Obesity describes health capital, in that it may make the body less productive, more susceptible to disease, and possibly cause it to depreciate more quickly. We will therefore look to see what part of the health capital model may explain it. We then consider some of its economic effects and finish with economic explanations as to why it has increased.

An Economic Treatment of Obesity

This discussion closely follows a model derived by Yaniv, Rosin, and Tobol (2009). They note that the human body needs energy to function, with food being the fuel that creates this energy. Potential energy exists in the form of calories burned in the process of daily functioning, and the body accumulates unburned energy in the form of fat tissues that increase body weight. People will gain or lose weight depending on the relationship of total calories consumed to total calories expended.

The body expends calories both in physical activity and at rest. The rest component, known as Basal Metabolic Rate (BMR), is the largest source of energy expenditure, reflecting blood circulation, respiration, and daily maintenance of body temperature. While the BMR is determined by physical characteristics (such as gender, age, weight, and height), calories expended through physical activity, as well as calorie intake through food

consumption, are subject to choice. Differing BMRs among individuals indicate why one person can "eat like a horse" and gain little weight, while another may gain weight with far less intake of food.

This economic theory of obesity views weight gain as the outcome of rational choice that reflects a willingness to trade off some future health for the present pleasures of less restrained eating and lower physical activity. Although economic models of obesity usually focus on food consumption in general as the source of energy, foods vary in their calorie content. Junk food, for example, is relatively high in calories, or *energy-dense*, while healthy food is lower in calories.

Suppose that overweight individuals can determine consumption of junk food meals, F, and healthy meals, H. They may also choose their level of exercise, x. The model defines the weight gain during a period, or obesity, S, as:

$$S = \delta F + \varepsilon H - \mu x - BMR \tag{7.5}$$

where δ and ε represent calorie intake per junk food meal F and healthy meal H, respectively (with $\delta > \varepsilon$), and μ represents calorie expenditure per instant of physical activity. In plain terms, people who eat too much and do not exercise enough will get fat.

Yaniv and colleagues note that people may eat even when they are not hungry, in social or stressful situations, and this type of eating may be composed of snacks, which are high in calories (i.e., part of F). Using F^S and M to denote snacks and hunger-induced meals, suppose that people can satisfy their hunger through either junk food F or healthy meals H. Hence,

$$\text{Meals} + \text{Snacks} = M + F^S = F + H \tag{7.6}$$

Substituting equation (7.6) into (7.5), the obesity function becomes

$$S = \varepsilon M + (\delta - \varepsilon)F + \varepsilon F^S - \mu x - BMR \tag{7.7}$$

If people satisfy hunger with healthy meals and healthy snacks alone so $F = 0$, then

$$S = \varepsilon(M + F^S) - \mu x - BMR \tag{7.7'}$$

To the extent that they substitute a junk food meal for a healthy meal or a snack, the value of S increases by $(\delta - \varepsilon)$.

Here, healthy food H does not enter the obesity function explicitly, but it moderates the calorie contribution of junk food meals that substitute in satisfying hunger and lead to increased weight. In this framework, taxes on junk food (reducing its consumption) or subsidies to healthy food (increasing its consumption) could have important impacts on the formation of health capital.

Economic Effects

This model provides several useful inferences about obesity. Clearly, body weight is a measure of health capital, and most people have at least some say in what happens to their level of health capital. Healthy eating (consuming H rather than F) exercising more (increasing exercise level x), and appropriate diet practices will help maintain appropriate body weight.

From a health investment point of view, obesity is a bad investment, leading to both higher medical expenditures, and lower earnings. Cawley and colleagues (2015) find that in 2010 obesity raised annual medical care costs by $3,508. Predicted medical care costs for both men and women just over the threshold of obesity are not significantly higher

than those of healthy weight individuals. However, costs rise exponentially in the morbidly obese range (BMI ≥ 35), which indicates that the greatest potential savings may come from facilitating weight loss in that group or preventing additional weight gain among obese individuals.

Cawley (2004) addresses the measured impact of obesity on wages. He cites several previous studies that found negative correlations between body weight and wages among females, with three broad explanations for this finding. First, obesity reduces wages by lowering productivity or because employers discriminate against obese people, paying them less than others. A second explanation is that low wages cause obesity. This may occur if poorer people consume cheaper, more fattening foods. The third explanation is that unobserved variables such as poor education cause both obesity and low wages.

In careful statistical analysis, he uses ordinary least squares (OLS) results to find that heavier white females, black females, Hispanic females, and Hispanic males tend to earn less, and heavier black males tend to earn more, than their less heavy counterparts. With more complex statistical models he shows that the effect is particularly strong for white females. A difference in weight of two standard deviations (roughly 64 pounds) is associated with a wage difference of 9 percent. The magnitude of this difference is equivalent in absolute value to the wage effect of roughly 1.5 years of education or three years of work experience.

Brunello and D'Hombres (2007) examine nine countries in the European Community Household Panel. While some of their country effects are not significant, due to small sample sizes, when they pool the data on all nine countries, they find that a 10 percent increase in BMI reduces the wages of females by 1.86 percent and males by 3.27 percent.

Why Has Obesity Increased?

Economists often seek to explain behaviors in terms of incomes and prices. Certainly for those who are abjectly poor and have too little to eat, increased incomes will increase weight. For income to explain obesity, one would have to show that as incomes grow, people tend to spend relatively more on energy-dense foods (those foods with a high number of calories per unit weight). Such evidence is generally lacking.

There is some evidence that the prices of energy-dense foods have fallen relative to others. Cawley (2015) reports findings that from 1990 to 2007, the real price of a 2-liter bottle of Coca-Cola fell 34.9 percent and the real price of a 12-inch Pizza Hut pizza fell 17.2 percent. In contrast the prices of fresh fruits and vegetables rose faster than inflation. If these food types are normal goods one might expect a shift in consumption to these more energy-dense foods.

Cutler, Glaeser, and Shapiro (2003) show that there was increased caloric intake for both men and women from the late 1970s to the late 1990s. Caloric intake increased for both men (almost 13 percent) and women (9.4 percent) over a 20-year period.

While there are multiple causes, the authors attribute substantial explanatory power to changes in the *time costs* of food production. Technological innovations—including vacuum packing, improved preservatives, deep freezing, artificial flavors, and microwaves—have enabled food manufacturers to cook food centrally and ship it to consumers for rapid consumption.

Table 7.4 shows that in 1965, married women who did not work outside the home spent over almost 138 minutes per day cooking and cleaning up from meals. In 1995, the same tasks took just about half the time. The switch from individual to mass preparation lowered the time price of food consumption and led to increased quantity and variety of foods consumed.

Table 7.4 Time Costs by Demographic Group (Minutes)

	1965		*1995*	
	Meal Prep.	*Meal Prep. + Cleanup*	*Meal Prep.*	*Meal Prep. + Cleanup*
Adults				
Single male	13.6	18.1	15.5	17.3
Married male, nonworking spouse	6.5	9.4	13.2	14.4
Married male, working spouse	8.1	11.9	13.2	14.4
Single female	38.1	60.1	28.9	33.1
Married female, working	58.3	84.8	35.7	41.4
Married female, not working	94.2	137.7	57.7	68.8
Elderly				
Male	16.6	26.3	18.5	20.2
Female	65.9	104.4	50.1	60.3

Source: Authors' calculations from Americans' Use of Time Survey Archives, 1965 and 1995. Permission AEA Publications.

They argue further, that peeling and cutting french fries, for example, is a marginal time cost, while deep frying (in restaurants) is generally a fixed cost (up to the point where the fryer is full). Mass preparation allows a restaurant to share the fixed time component over a wide range of consumers. In addition, mass preparation reduces the marginal cost of preparing food by substituting capital for labor. Finally, mass preparation exploits the division of labor. Food professionals instead of "home producers" now prepare food, reducing both fixed and marginal costs.

The authors calculate that reductions in the time required to prepare food reduced the per-calorie cost of food by 29 percent from 1965 to 1995. If the elasticity of caloric intake with respect to price is –0.7, this could explain the increase in caloric intake and the corresponding increases in obesity. If the calorie intake elasticity is a bit less responsive (say –0.5), as is likely, then issues of individual self-control, food advertising, and perhaps lack of information on the true costs of obesity may also explain the serious increase in obesity over the past 20 to 30 years.

Conclusions

This chapter has addressed the demand for health and medical services from an individual perspective. It has treated health as a good produced like all others, using market inputs as well as an individual's time. People benefit from health in four important ways:

1 They feel better when well.
2 They lose less time to illness, and hence can work more.
3 They are more productive when they work and can earn more for each hour they work.
4 They may live longer.

By analyzing the demand for health in this way, we recognize that the demands for health care inputs—from physicians' services, to drugs, to therapy—are demands that are derived from the demand for health itself. Consumers, jointly with providers, allocate resources among health care inputs based on the demand for health. We address the specific demand for health inputs in Chapter 9.

Summary

1 It is not medical care as such that consumers want, but rather health itself. Medical care demand is a derived demand for an input that is used to produce health.
2 Consumers do not merely purchase health passively from the market, but instead produce it, spending time on health-improving efforts in addition to purchasing medical inputs.
3 Health lasts for more than one period. It does not depreciate instantly, and thus it can be treated as a capital good.
4 Demand for health has pure consumption aspects; health is desired because it makes people feel better.
5 Demand for health also has pure investment aspects; health is desired because it increases the number of healthy days available to work and thus earn income.
6 Consumers prefer more income and more leisure so indifference curves between income and leisure are negatively sloped. The slope of the line depicting the labor–leisure trade-off is the after-tax wage rate.
7 Because health is a capital good, the cost of holding health for any one year, as well as over time, will equal the opportunity cost of the capital (interest foregone) plus the depreciation (deterioration of value).
8 The *MEI* curve for investments in health is downward sloping because the production function for healthy days exhibits diminishing marginal returns. Thus, the optimal demand for health is likewise given at the intersection of the *MEI* curve and the cost of capital curve $(r + \delta)$.
9 The pure investment model generates the prediction that optimal health will decline as the person ages, if the depreciation rate of health increases as a person ages.
10 The rewards of being healthy are generally greater for higher-wage workers so those with increasing wages will generally tend to increase their optimal health stock.
11 Health can be generated at less cost for more highly educated people, resulting in a higher optimal health stock than for less educated people.
12 Obesity is a prime example of health capital analysis. Many feel that decreased time and money costs of energy-dense foods, and food preparation, have had substantial impacts on caloric intake, and hence obesity.

Discussion Questions

1 Why do we treat leisure and earnings as ordinary utility-increasing goods?
2 Describe the aspects of health that make it a consumption good. Describe those that make it an investment good.
3 Give examples of how health is produced from time and market goods.
4 Why is the depreciation of a capital good a cost to society? In what ways does a person's health depreciate?

5 Why might older people's health care expenditures increase in the Grossman model even though their desired health stock may be lower?
6 List at least three factors that might increase an individual's marginal efficiency of investment in health capital.
7 Suppose that a young woman goes on to medical school and becomes a physician. Would you expect her expenditures on medical goods for her own health to be higher or lower than a nonphysician? Why?
8 From your experience, do you think the typical person becomes less healthy upon, or shortly after, retirement? What does the Grossman model predict?
9 People who earn a higher salary can afford more goods, including health care. However, according to Grossman, they will choose a higher desired health stock. Why is this so, according to the model?
10 Knowing the potential negative effects, would a "rational" person ever choose to become obese.

Exercises

1 Draw an isoquant (see Chapters 2 and 6) for medical inputs and other inputs in the production of a given amount of health investment. What does the isoquant mean? How would the isoquant look if substitution was limited? If a high degree of substitution was possible?
2 Suppose that no amount of other goods can compensate for a loss in health. How would the individual's indifference curves look? Is this a reasonable assumption in terms of what we actually see taking place?
3 Suppose that John Smith gets promoted to a job that causes two changes to occur simultaneously: John earns a higher wage, and a safer environment causes his health to depreciate less rapidly. How would these two changes together affect John's desired health capital?
4 Suppose that John could work 365 days per year and could earn $200 per day for each day he worked. Draw his budget line with respect to his labor–leisure choice.
5 Suppose that John chooses to work 200 days per year. Draw the appropriate indifference curve, and note his equilibrium wage income and labor–leisure choices.
6 Suppose, in Exercise 5, that John's wage rises from $200 to $210 per day. Show how his equilibrium level of income and labor–leisure will change.
7 Suppose that John is ill 10 days per year. Draw the impact of this illness on the equilibrium defined in Exercise 5. How will it change his equilibrium allocation of earnings and labor vs. leisure?
8 Answer the following.
   (a) Depict Sara's optimal stock of health capital at age 18, with a high school diploma and a wage of $10 per hour.
   (b) Suppose that she invests in a college education, expecting to get a better and higher-wage job. Show how her optimal stock of health capital changes by the age of 30 due to the increased wage. Then, show how her education would affect her optimal health stock if education also made her a more efficient producer of health.
   (c) Suppose that after age 30 her wage stays the same. As she ages, show what happens to her optimal stock of health capital, assuming that the depreciation rate of health increases with age.

9 Consider Fred's investment in units of health capital with the following function:

$$I = 500 - 1{,}000 \times \text{cost of capital.}$$

(a) Indicate some of the components of the cost of capital, and why they are costs.
(b) If the cost of capital is 10 percent each year, what is the equilibrium health investment in terms of units of capital? What is the equilibrium total investment expenditure? Explain both answers.
(c) If the cost per unit of health capital doubles to 20 percent, what will happen to the equilibrium level of health investment and to equilibrium health investment expenditures? Why?

10 Consider the obesity model where equation (7.7) refers to a daily obesity function. Suppose that Ed can either exercise 0 or 1 session per day, with $\mu = 300$. If Ed substitutes one junk food meal ($\delta = 800$) for a healthy meal ($\varepsilon = 600$) five days per week, how often per week will he have to exercise to avoid increasing S? Why?

Notes

1 We ignore here income from nonwork efforts—for example, through returns to financial investments, such as saving, stock, or bonds.
2 Goodman, Stano, and Tilford (1999) provide a more detailed model addressing the production of health and home goods using both market goods, and people's time.
3 This result *may* be ambiguous. Although an increased wage rate potentially increases the return to investment, it also represents an increased opportunity cost of time in producing health investment. If health investment has a large labor component, and wage rises, the *MEI* curve could shift downward, and it is possible that the equilibrium demand for health investment will fall.

Chapter 8

Demand and Supply of Health Insurance

In this chapter

- What Is Insurance?
- Risk and Insurance
- The Demand for Insurance
- The Supply of Insurance
- The Case of Moral Hazard
- Health Insurance and the Efficient Allocation of Resources
- Income Transfer Effects of Insurance
- Conclusions

Health insurance underlies any discussion of the health economy. Most Americans, and indeed most citizens of other countries, do not pay directly for their health care. Rather, private or public insurers pay for much of the care, with the consumer paying only a portion of the bill directly. This portion is sometimes called coinsurance. Insurance coverage is provided through the payment of premiums (in privately financed systems) or taxes (when insurance is provided publicly). In the United States, the premiums have often, although not always, been purchased and paid for through the consumer's participation in the labor force.

Health care expenses are uncertain because many illnesses occur rarely and seemingly at random. When they do, they may cost a great deal, and they can be financially troublesome if not ruinous to households. Costs could be so high that without financial help, treatment might not be available.

Because insurance is so important to the demand and supply of health care, as well as the government's role in allocating health resources, we spend this entire chapter on the demand and supply of insurance in general, and health insurance in particular. This provides a set of tools for addressing issues such as the demand for and supply of health care, the role of information in health care markets, and the variation of health care among various markets. We return to insurance issues in Chapter 11, which will look at the operation of insurance markets.

What Is Insurance?

Consider the demand for insurance without all of the detailed trappings (deductibles, premiums, coinsurance, etc.) that accompany modern insurance plans. Start with a club with 100 members. The members are about the same age, and they have about the same interests and lifestyles. About once a year one of the 100 members gets sick and incurs health care costs of $5,000. The incidence of illness seems to be random, not necessarily striking men, women, the old, or the young in any systematic fashion. Club members, worried about potential financial losses due to illness, decide to collect $50 from each member and put the $5,000 in the bank for safekeeping and to earn a little interest. If a member becomes ill, the fund pays for the treatment. This, in a nutshell, is insurance. The members have paid $50 in advance to avoid the risk or uncertainty, however small, of having to pay $5,000. The "insurer" collects the money, tries to maintain and/or increase its value through investment, and pays claims when asked.

This example illustrates several desirable characteristics of an insurance arrangement.

1 The number of insured should be large, and they should be independently exposed to the potential loss.
2 The losses covered should be definite in time, place, and amount.
3 The chance of loss should be measurable.
4 The loss should be accidental from the viewpoint of the person who is insured.

Insurance generally reduces the variability of the incomes of those insured by pooling a large number of people and operating on the principle of the *law of large numbers*. That is, although outlays for a health event may be highly variable for any given person in the insurance pool, insurers can predict the average outlays for the group. The law of large numbers shows that for a given probability of illness, the distribution of the average rate of illness in the group will collapse around the probability of illness as the group size increases.

This chapter considers the theory and practice of health care insurance. It shows the necessity of quantifying risk, as well as attitudes toward risk. With those ideas, we consider the structure of insurance policies and how markets evolve to provide them.

Insurance versus Social Insurance

Insurance is provided through markets in which buyers protect themselves against rare events with probabilities that can be estimated statistically. The government programs use the government as insurer and are distinguished by two features:

1 Premiums (the amounts paid by purchasers) are heavily and often completely (as in the case of Medicaid) subsidized.
2 Participation is constrained according to government-set eligibility rules.

In addition, government insurance programs often transfer income from one segment of society to another. Given the importance of such social insurance programs, we devote an entire chapter (Chapter 20) to them later in the text.

Insurance Terminology

Consider some terms that we use to discuss insurance. Although much of the analysis uses the standard economic language of prices and quantities, the insurance industry has developed a particular set of definitions. These include:

Premium, Coverage—When people buy insurance policies, they typically pay premiums for a given amount of coverage should the event occur. For example, an insured person may pay a $50 premium for $1,000 of coverage.

Coinsurance and Copayment—Many insurance policies, particularly in the health insurance industry, require that when events occur, the insured person share the loss through copayments. This percentage paid by the insured person is the coinsurance rate. With a 20 percent coinsurance rate, an insured person, for example, would be liable (out-of-pocket) for a $30 copayment out of a $150 charge. The insurance company would pay the $120 remainder, or 80 percent. Thus, coinsurance refers to the *percentage* paid by the insured; copayment refers to the *amount* paid by the insured (such as a fixed payment for a prescription).

Deductible—With many policies, the insured must pay some amount of the health care cost in the form of a deductible, irrespective of coinsurance. In a sense, the insurance does not apply until the consumer pays the deductible. Deductibles may apply toward individual claims. Often in the case of health insurance they apply only to a certain amount of total charges in any given year.

Insurers often use coinsurance and deductibles together. An insurance policy may require that Elizabeth pay the first $250 of her medical expenses out-of-pocket each year. It may then require that she pay 20 percent of each additional dollar in charges. This policy then has a deductible of $250 and a coinsurance rate of 20 percent.

Many feel that deductibles and coinsurance simply represent ways that insurance companies have found to separate consumers from their money. Economists, in contrast, have explained that deductibles and coinsurance may lead to desirable economic consequences. The copayment requirement makes consumers more alert to differences in the true costs of the treatment they are purchasing. Charging deductibles discourages frivolous claims or visits, and it also makes insured people more aware of the results of their actions. Both deductibles and coinsurance may serve to avoid claims and to reduce costs.

Finally, a few other terms describe various features of insurance:

Exclusions—Services or conditions not covered by the insurance policy, such as cosmetic or experimental treatments.

Limitations—Maximum coverages provided by insurance policies. For example, a policy may provide a maximum of $3 million lifetime coverage.

Pre-Existing Conditions—Medical problems not covered if the problems existed prior to issuance of insurance policy. Examples here might include pregnancy, cancer, or HIV/AIDS. The Affordable Care Act (ACA) of 2010 prohibits the denial of insurance coverage due to pre-existing conditions.

Pure Premiums—The actuarial losses associated with the events being insured.

Loading Fees—General costs associated with the insurance company doing business, such as sales, advertising, or profit.

With these in mind, we turn to a more formal analysis of risk and insurance.

Risk and Insurance

To this point we have assumed that all decisions occur under conditions of certainty; that is, consumers know what the prices, incomes, and tastes are and will be. Clearly, however, many decisions are made under conditions in which the outcome is risky or uncertain. Students should know that economists sometimes contrast risk, where the probability of an adverse effect is known (like the odds of a roulette wheel), with uncertainty, where the probability is not known (the odds of a nuclear plant meltdown). Our discussion will use the terms interchangeably.

We begin by considering the insurance coverage of an event that occurs with the known probability, p, leading to a predictable loss and/or payment. This assumption will characterize people's choices under uncertainty. We will then extend the general characterization to health insurance, where the payment may be affected by the insurance. We address this difference once the basic points regarding risk are developed.

Expected Value

Suppose Elizabeth considers playing a "coin flip" game. If the coin comes up heads, Elizabeth will win $1; if it comes up tails, she will win nothing. How much would Elizabeth be willing to pay in order to play this game? Analysts rely on the concept of expected value for the answer. With an honest coin, the probability of heads is one-half (0.5), as is the probability of tails. The expected value, sometimes called the expected return, is:

$$E(\text{heads is called}) = (\text{probability of heads}) \times (\text{return if heads, i.e., } \$1) + (\text{probability of tails}) \times (\text{return if tails, } \$0) \quad (8.1)$$

The expected value is $0.50; that is (½ × 1) + (½ × 0). If she uses the decision criterion that she will play the game if the expected return exceeds the expected cost, Elizabeth will play (pass) if it costs her less (more) than $0.50. More generally, with n outcomes, expected value E is written as:

$$E = p_1R_1 + p_2R_2 + \ldots + p_nR_n$$

where p_i is the probability of outcome i (that is p_1 or p_2, through p_n) and R_i is the return if outcome i occurs. The sum of the probabilities p_i equals 1.

The special case where the price of the gamble is exactly $0.50 and equals the expected return is analogous to an insurance situation in which the expected benefits paid out by the insurance company equal the premiums taken in. This equality of expected benefit payments

and premiums is called an *actuarially fair* insurance policy. In reality, insurance companies must also cover additional administration and transaction costs to break even (the loading costs discussed earlier), but the definition of an actuarially fair policy provides a benchmark in talking about insurance.

Marginal Utility of Wealth and Risk Aversion

The foregoing example implies that Elizabeth is indifferent to risk. That is, her incremental pleasure of winning \$0.50 (the gain of \$1 less the \$0.50 she paid to play) is exactly balanced by her incremental displeasure of losing \$0.50 (the gain of zero less the \$0.50 paid to play). Suppose we now increase the bets so that the coin flip now yields \$100, or nothing, and that Elizabeth is now asked to bet \$50 to play. Actuarially this is the same bet as before, but Elizabeth may now think a little harder. She may now refuse an actuarially fair bet—\$50—on the grounds that she cannot afford to risk the \$50 loss if the coin lands tails. This suggests that the disutility of losing \$50 may exceed the utility of winning \$50. This would occur if she felt that the utility of an extra dollar of wealth is greater if she has less money than the utility of an extra dollar of wealth when she has more. The utility from an extra dollar is called the marginal utility of wealth.[1]

It is important to incorporate Elizabeth's utility of wealth function into the analysis. In Chapter 2, we assumed that consumers could rank bundles but could not (and need not) compare magnitudes of satisfaction. Here, however, to understand the utility model of risk behavior, we must further assume that consumers can rank alternatives and compare their magnitudes.

In Figure 8.1 suppose that Elizabeth's wealth is \$10,000. That wealth gives her a utility level of $U_1 = 140$ and allows her to buy some basic necessities of life. This can be denoted as point *A*. Suppose her wealth rises to \$20,000. Will her utility double?

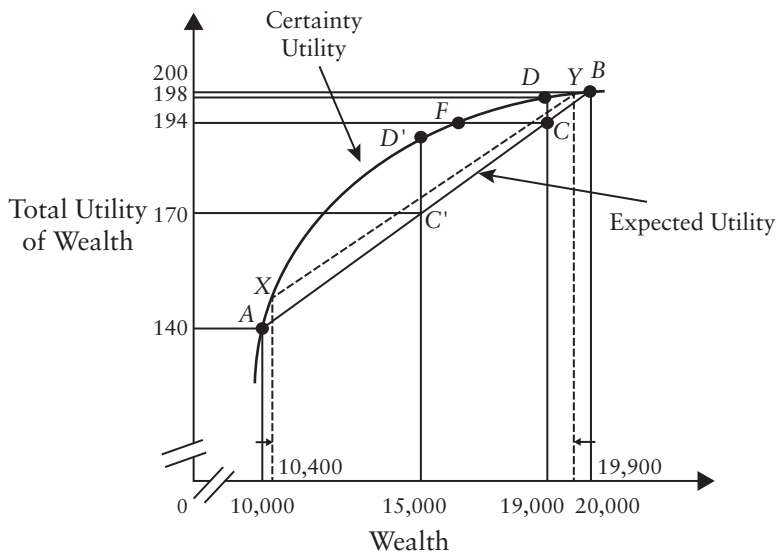


Figure 8.1 Total Utility of Wealth and the Impact of Insurance

While it is hard to be certain, most likely the next $10,000 will not bring her the incremental utility that the first $10,000 brought. So, while U_2 is certainly higher than U_1, it (U_2) will likely be less than twice U_1. Suppose, for example, that $U_2 = 200$. This is denoted as point *B*. Do all of the points on the utility function between U_1 and U_2 lie on a straight line? If they do, this is equivalent to saying that the utility from the 10,001st dollar is equal to the utility from the 19,999th dollar, and hence the marginal utility is constant. This also is unlikely. Because the marginal utility of earlier dollars is likely to be larger than that of later dollars, the utility curve is likely to be bowed out, or concave, to the *x*-axis.

The marginal utility of wealth refers to the amount by which utility increases when wealth goes up by $1. This rise in utility, divided by the $1 increase in wealth, is thus the slope of the utility function. The bowed shape of the utility function shows a slope that is getting smaller or flatter as wealth rises; the marginal utility of wealth is diminishing.

Elizabeth begins with wealth of $20,000, but understands that if she falls ill, which may occur with probability 0.10, the expenses will cause her wealth to decline to $10,000. If this occurs, she can calculate her expected wealth, $E(W)$,

$$E(W) = (\text{prob. well} \times \textbf{wealth} \text{ if well}) + (\text{prob. ill} \times \textbf{wealth} \text{ if ill})$$
$$(0.90 \times \$20{,}000) + (0.10 \times \$10{,}000) = \$19{,}000 \qquad (8.2a)$$

and expected utility, $E(U)$:

$$E(U) = (\text{prob. well} \times \textbf{utility} \text{ if well}) + (\text{prob. ill} \times \textbf{utility} \text{ if ill})$$
$$(0.90 \times \text{utility of } \$20{,}000) + (0.10 \times \text{utility of } \$10{,}000)$$
$$(0.90 \times 200) + (0.10 \times 140) = 194 \qquad (8.2b)$$

Thus, the expected utility $E(U)$ is 194 or point *C* because of the risk of illness. Geometrically, this is the line segment between points *A* and *B*, evaluated at wealth level $E(W) = \$19{,}000$. The expected utility due to risk must be compared to the utility of 198 (point *D*), corresponding to the utility that she would receive if she could purchase insurance at an actuarially fair rate. As drawn, the *risk of loss* puts her on the line below the curve indicating certainty, and leads to a loss of 4 units (198 to 194) of utility. To be clear, Elizabeth worries that her wealth may be either $10,000 or $20,000, and this uncertainty costs her utility. She would have higher utility if she could eliminate this uncertainty.

Purchasing Insurance

Suppose that Elizabeth can buy an insurance policy with a premium of $1,000 per year that will maintain her wealth irrespective of her health, thus eliminating the uncertainty. That is, if she stays well, her wealth will be $20,000 less the $1,000 premium. If she falls ill, she is provided $10,000 in benefits, so that her wealth will be $10,000 plus the $10,000 in benefits, less the $1,000 premium. She is certain to have $19,000 at the end of the year.

Is it a good buy? At a net wealth of $19,000, which equals her initial wealth minus the insurance premium, Elizabeth's certainty utility is 198. She is better off at point *D* than at point *C*, as shown by the fact that point *D* gives the higher utility. If insuring to get a *certain* wealth rather than facing the risky prospect makes Elizabeth better off, she will insure.

We can, in fact, use Figure 8.1 to calculate the maximum amount that Elizabeth would be willing to pay for the insurance by moving southwest down the utility function to the level of $U = 194$ and reading the level of wealth (off the *x*-axis) to which it corresponds, or point *F*. The distance *FC* reflects Elizabeth's aversion to risk. At point *F*, Elizabeth would be willing to pay up to $4,000 (that is, initial wealth of $20,000, less $16,000 at point *F*) for insurance

and still be as well off as if she had remained uninsured. If, for example, she were able to purchase the insurance for $3,000, she would get $1,000 in consumer surplus.

This analysis illustrates several facts:

1 Consumers will buy insurance only when there is diminishing marginal utility of wealth or income—that is, when the consumer is risk-averse. In Figure 8.1, if marginal utility is constant, a requirement that Elizabeth pay an actuarially fair premium for insurance would leave her no better off than if she were uninsured. If Elizabeth is "risk-loving," with increasing marginal utility of wealth or income, she will definitely refuse to buy insurance.
2 Expected utility is an average measure; Elizabeth either wins or loses the bet. If exposed to risk, Elizabeth will have wealth and hence utility of either $20,000 (with utility of 200), or $10,000 (with utility of 140), and a risky expected wealth of $19,000. Insurance will guarantee her wealth to be $19,000. If she stays well, her wealth will be $20,000 less the $1,000 insurance premium; if she falls ill, her wealth will be $10,000 plus the $10,000 payment for the loss of health, minus the $1,000 premium—again $19,000.
3 If insurance companies charge more than the actuarially fair premium, people will have less expected wealth from insuring than from not insuring. Even though they will have less wealth as a result of their insurance purchases, the increased well-being comes from the elimination of risk.
4 The willingness to buy insurance is related to the distance between the utility curve and the expected utility line. If Elizabeth is very unlikely to become ill (near point *B*), then her expected utility will be almost identical to her certainty utility, and her gains from insurance will be small. If Elizabeth's probability of illness increases to 0.5 (point *C´*), her expected wealth will be $15,000 and her expected utility will be 170. She will accrue significant gains by insuring as noted by the distance *C´D´*. However, if Elizabeth is almost certain to fall ill, (approaching point *A*), her gains from buying insurance decrease. Why? Here, she is better off "self-insuring," by putting the (almost) $10,000 away to pay for her almost certain illness rather than incurring the trouble of buying insurance and then filing claims.

The Demand for Insurance

Exactly how much insurance will Elizabeth purchase? The next two sections present a classic model introduced by Mark Pauly in 1968 to consider the fundamental demand and supply decisions regarding insurance.

How Much Insurance?

Recall that Elizabeth's expected utility involves her wealth when ill, with a probability of 0.10, or when healthy, with a probability of 0.90. If ill, her wealth will fall from $20,000 to $10,000.

We address Elizabeth's optimal purchase using marginal benefits and marginal costs. Consider first a policy that provides insurance covering losses up to $500. Although Elizabeth might find it hard to justify buying a $500 insurance policy when she will lose $10,000 if she falls ill, it is a useful place to start.

The goal of maximizing total net benefits provides the framework for understanding her health insurance choice. She benefits from health insurance only when she is ill and receives the insurance benefit payments. She still pays the insurance premiums when ill, but gains financially net of those premiums. When well only the premium applies to her, and this is a net cost.

In turn when she is well, the marginal benefits will decline as she purchases additional insurance coverage; we understand this by applying the law of diminishing marginal utility of wealth. In contrast, again due to the diminishing marginal utility of wealth, the marginal costs when well will rise as when she purchases additional insurance. At the end, Elizabeth will buy insurance so that the marginal benefits of the last dollar spent equal the marginal costs.

Suppose she must pay a 20 percent premium ($100) for her insurance, or $2 for every $10 of coverage that she purchases. The following worksheet describes her wealth if she gets sick.

Insurance Worksheet—$500 Coverage
Wealth If Ill

	Original wealth	$20,000
less	Loss	$10,000
	Remainder	$10,000
plus	Insurance	500
	Sum	$10,500
less	Premium	100
or	**New** wealth	$10,400

For the initial coverage, Elizabeth's wealth if well is $20,000 less the $100 premium, or $19,900. Her marginal benefit from the $500 from insurance is the expected marginal utility that the additional $400 ($500 minus the $100 premium) brings. Her marginal cost is the expected marginal utility that the $100 premium costs. We describe these benefits and costs in Figure 8.2. If Elizabeth is averse to risk, the marginal benefit (point *A*) of this insurance policy exceeds its marginal cost (point *A´*).

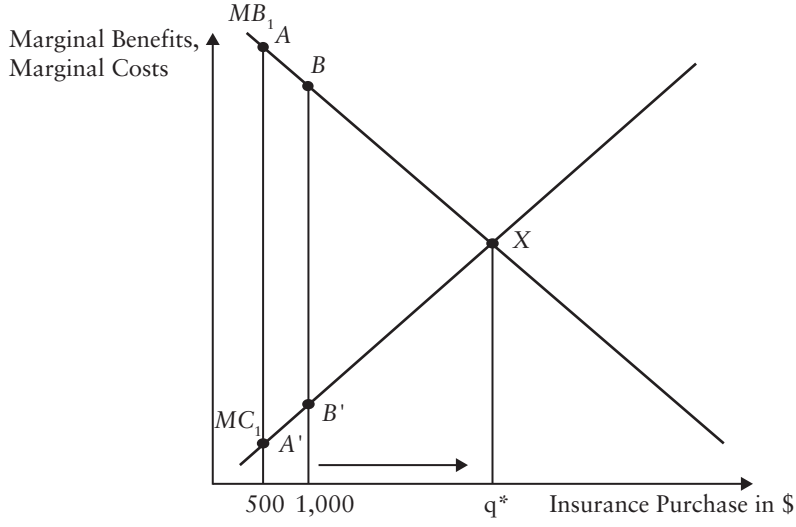


Figure 8.2 The Optimal Amount of Insurance

We can also see this in the original Figure 8.1. Point *X* refers to Elizabeth's wealth if ill, and point *Y* to her wealth if healthy. By inspection, we see that the incremental (marginal) utility between points *A* and *X* increases much more than it decreases between points *B* and *Y*. Expected utility line *XY* lies above the original expected utility line *AB*. The first $500 of insurance is a good buy!

Should Elizabeth increase her coverage from $500 to $1,000? She must again compare the marginal benefits of this next $500 increment to its marginal costs. Because Elizabeth is slightly wealthier than before, if ill (starting at $10,400 rather than $10,000) the marginal utility from an additional $400 of wealth (calculated as before) will be slightly smaller than from the first $400. Hence, the marginal benefits from the second $500 insurance increment will be slightly smaller than for the first $500 increment. Her marginal benefit curve, MB_1, is downward sloping, with her new marginal benefit at point *B*.

Similarly, because if well she is a little less wealthy than before, an additional $100 in premiums will cost a little more in foregone (marginal) utility of wealth than the first increment at point *B´*. Thus, her marginal cost curve, MC_1, is upward sloping.

Continuing, we see that Elizabeth will adjust amount of insurance *q* that she purchases to the point at which the marginal benefits equal the marginal cost. The quantity, q^*, at which they are equal (point *X*) is Elizabeth's optimum insurance purchase. The *x*-axis of Figure 8.2 is drawn to scale, and it shows that q^* is approximately $3,000.

Changes in Premiums

How will her insurance decision change if premiums change, that is insurers raise the prices for the product they sell? Consider first the impact of a higher premium, say 25 percent rather than the 20 percent used earlier. With the 25 percent premium ($125), Elizabeth faces the following calculation for the starting $500 policy:

Insurance Worksheet—Higher Premium		
Wealth If Ill		
	Original wealth	$20,000
less	Loss	$10,000
	Remainder	$10,000
plus	Insurance	500
	Sum	$10,500
less	*New premium*	125
or	**New** wealth	$10,375

If she stays well, her wealth is $20,000 less the $125 premium, or $19,875. Look now at Figure 8.3. Elizabeth's marginal benefit from the $500 from insurance is now $375 rather than the previous value of $400, so point *C* lies on curve MB_2 below the previous marginal benefit curve, MB_1. We can fill in additional points on this curve, which reflects the higher premium.

Similarly, Elizabeth's marginal cost is the expected marginal utility that the (new) $125 premium costs her. This exceeds the previous cost in terms of foregone utility, so point *C* lies

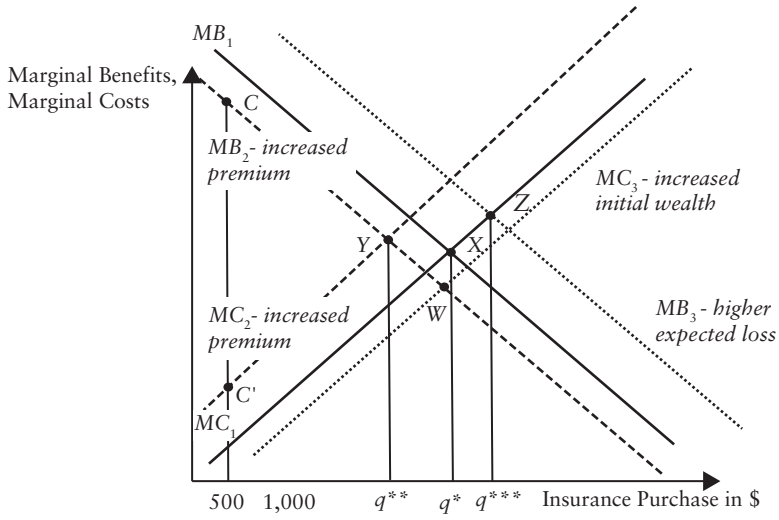


Figure 8.3 Changes in the Optimal Amount of Insurance

on curve MC_2 above the previous marginal cost curve, MC_1. Again, we can fill in additional points on this curve and find the intersection of MB_2 and MC_2 at point *Y*. The resulting analysis suggests that consumers react rationally to higher premiums by reducing their optimum coverage from q^* to q^{**}. In this example, the purchase fall is from about $3,000 to $2,300.

Changes in Expected Loss

How will changes in expected losses affect the insurance decision? Returning to the original example with a premium of 20 percent, suppose that instead of $10,000, Elizabeth expected to lose $15,000 if ill. Consider again the first $500 of insurance coverage. Her wealth, if healthy, is $19,900, so nothing changes with respect to marginal cost. Elizabeth remains on curve MC_1. The marginal benefit calculation, however, does change:

Insurance Worksheet—Higher Expected Loss

Wealth If Ill

	Original wealth	$20,000
less	*New loss*	$15,000
	Remainder	$5,000
plus	Insurance	500
	Sum	$5,500
less	Premium	100
or	**New** wealth	$5,400

As before, the insurance gives her a net benefit of $400. However, this net benefit increments a wealth of $5,000 rather than $10,000. If we assume that an additional dollar gives more marginal benefit from a base of $5,000 than from a base of $10,000, then the marginal benefit curve shifts upward because of the increased expected loss. This provides equilibrium point Z on curve MB_3 in Figure 8.3. It follows that in equilibrium, an increase in the expected loss will increase the amount of insurance purchased at point Z, or q^{***}. As drawn, q^{***} equals approximately $3,500.

Changes in Wealth

Finally, consider a change in initial wealth. Suppose Elizabeth started with a wealth of $25,000 instead of $20,000. Assume once again a premium rate of 20 percent.

Insurance Worksheet—Increased Wealth
Wealth If Ill

	Increased wealth	$25,000
less	Loss	$10,000
	Remainder	$15,000
plus	Insurance	500
	Sum	$15,500
less	Premium	100
or	**New** wealth	$15,400

At the higher level of wealth, the same insurance policy provides a smaller increment in utility, so the marginal benefit curve shifts down from MB_1 to MB_2. However (for the same expected loss), the $100 premium costs less in foregone marginal utility relative to the increased wealth, a downward shift of MC_1 to MC_3. As a result of both downward shifts, the new equilibrium value of q at point W may be higher or lower than the original value of q^* (as drawn in Figure 8.3, it is slightly lower, about $2,800). If, however, increased wealth is accompanied by increased losses, then the MB curve may shift down less. If it does shift down by less, the desired amount of insurance may increase because the increased expected losses would make a larger amount of insurance more desirable.

The Supply of Insurance

In the previous example, we *assumed* a 20 percent premium rate, but to determine the amount of coverage someone will buy, we must know how insurers determine the premium. We started this chapter with the club that insures its members against illness. The officers of the club do not know, nor necessarily care, who will file a claim.[2] To function as an insurer, the club must simply see that that revenues cover costs. In practice, insurers will also incur administrative and other expenses that also must be covered by premiums.

Competition and Normal Profits

Let's return to Elizabeth's insurance problem from the previous section. In a competitive market, and under perfect competition, all firms earn zero excess profits. Recall that Elizabeth faced a potential illness with a probability of 0.10 (1 in 10). She sought to buy insurance in blocks of $500, and at the outset, her insurer, Asteroid Insurance, was charging her $100 for each block of coverage, or an insurance premium of 20 percent ($100 as a fraction of $500). Assume also that it costs Asteroid $8 annually to process each insurance policy and (if necessary) write a check to cover a claim. Asteroid's profits per policy are:

$$\text{Profits} = \text{Total revenue} - \text{Total costs}$$

Revenues are $100 per $500 policy. What are Asteroid's costs? For 90 percent of the policies, costs are $8 because the insured does not get sick and does not collect insurance. The company pays only the $8 processing costs per policy. The costs for the other 10 percent of the policies costs are $508, consisting of the $500 payment to those who are ill plus the $8 processing costs.

Thus, the per policy profits for Asteroid are:

$$\begin{aligned} \text{Profits} &= \$100 - (\text{Probability of illness} \times \text{Costs if ill}) \\ &\quad - (\text{Probability of no illness} \times \text{Costs if no illness}) \\ \text{Profits} &= \$100 - (0.10 \times \$508) - (0.90 \times \$8) \\ \text{Profits} &= \$100 - \$50.80 - \$7.20) \end{aligned} \tag{8.3}$$

$$\text{Profits (premium} = 10\%) = \$100 \text{ (revenues)} - \$58 \text{ (costs)} = \$42$$

These are positive profits for Asteroid, and they imply that a competing firm, Comet Insurance, (also incurring costs of $8 to process each policy) might enter the market and charge a lower premium, say, 15 percent, to attract clients. The cost side of the equation would remain the same, but the revenues for the two competitors, which equal the premium fraction multiplied by the amount of insurance, would fall to $75. Hence, profits fall to:

$$\text{Profits (premium} = 15\%) = \$75 \text{ (revenues} - \$58 \text{ (costs)} = \$17$$

which is still positive. We can see that entry will continue into this industry until the premium has fallen to a little less than 11.6 cents per dollar of insurance, which would provide revenues of $58, offset by the $58 in costs, to give zero profits.

Some algebra will verify that the 11.6 percent premium is directly related to the probability of the claim (10 percent). Quite simply, for Asteroid, the revenue per policy is aq, where a is the premium, in fractional terms. Algebraically, the cost per policy in terms of payout is the probability of payout, p, multiplied by the amount of payout, q, plus a processing cost, t, which is unrelated to the size of the policy (assuming it costs no more to administer a $10,000 policy than a $500 policy). So:

$$\text{Profits} = \text{Revenue } (aq) - \text{Cost}(pq + t) = aq - (pq + t) = aq - pq - t \tag{8.4}$$

With perfect competition, profits must equal 0, so:

$$\text{Profits} = 0 = aq - pq - t$$

We solve for the competitive premium a as:

$$a = p + (t/q) \tag{8.5}$$

This expression shows that the competitive value of a equals the probability of illness, p, plus the processing (or loading) costs as a percentage of policy value, q, or t/q. If loading costs are 10 percent of the policy value, q, then $(t/q) = 0.10$. Hence, in equilibrium, if p equals 0.10, then $a = p + (t/q) = (0.10 + 0.10) = 0.20$. The premium for each dollar of insurance, q, is \$0.20. If insurers charge less, they will not have enough money to pay claims. If they charge more, firms like Asteroid will earn excess profits, and other firms (like Comet) will bid down rates in perfectly competitive markets.

Previously, in discussing risk bearing, we considered insurance policies that would compensate Elizabeth against the loss based solely on the probability of the event's occurring. We refer to such rates as *actuarially fair* rates. The actuarially fair rates correspond to the rates in which the loading costs t as a percentage of insurance coverage, q (that is, t/q), approach 0, hence:

$$a = p + (t/q) = p + 0 = p \tag{8.6}$$

Knowing that premium a equals p under perfect competition (with no loading costs), we now solve for the optimal coverage against any expected loss. To maximize utility, Elizabeth will add coverage up to the point where her expected wealth will be the same whether she is ill or well. In the earliest example, the particular illness occurred with a probability of 0.10 and incurred a loss of \$10,000. In a competitive insurance market (ignoring loading costs), Elizabeth's wealth, if well, will be:

Wealth (if well) = \$20,000 – cost of insurance, or:

$$\$20,000 - (\text{premium } a) \times (\text{coverage } q)$$

Her wealth, if ill, will be:

$$\text{Wealth (if ill)} = \$20,000 - \text{loss} + \text{insurance reimbursement} - \text{insurance premium or:}$$
$$\$20,000 - \$10,000 \text{ loss} + \text{coverage } q - (\text{premium } a) \times (\text{coverage } q)$$

To maximize the expected utility we equate the wealth if well with the wealth if ill:

$$[\textbf{Wealth (if well)} = \$20,000 - aq] = [\$20,000 - \$10,000 = q - aq = \textbf{Wealth (if ill)}] \tag{8.7}$$

Subtracting \$20,000 – aq from both sides and rearranging terms yields:

$$q^{opt} = \$10,000$$

Elizabeth's optimal level of coverage for a loss of \$10,000 in the absence of transactions costs is \$10,000, irrespective of the probability of the event! It is more realistic of course to assume that transactions costs t will be positive; in fact, loading charges are often substantial. Under these circumstances, Elizabeth's best choice is to insure for less than the full health expense we show in the examples above, where the transaction costs were positive and optimal insurance q^* was considerably smaller than \$10,000. This standard result from the mathematics of health insurance contrasts with the propensity of consumers to seek full coverage.

The Case of Moral Hazard

To this point, we have discussed the theory of risk, as well as the demand and supply of insurance when the events and the losses are random. The insurance policies discussed thus

far represent indemnity policies, in which the insurer's liability is determined by a fixed, predetermined amount for a covered event. Indeed, the term *indemnity* comes from the Latin *indemnis*, meaning "unhurt." Insurance renders the insured party financially unhurt by the random event.

In the previous section, we showed that the optimal insurance policy covers the entire loss when there is no transaction cost, and less than the full loss in the more realistic case when transactions or loading costs are positive. We now address the effects of the price system on the provision of insurance.

Our discussions have assumed a fixed loss—that did not change merely because people bought insurance. However, in many cases, buying insurance lowers the price per unit of service to consumers at the time that they are purchasing services. If people purchase more service due to insurance, then we must modify many of the insurance propositions just presented.

Demand for Care and Moral Hazard

Suppose Elizabeth faces the probability $1 - p = 0.5$ that she will not be sick during a given time period and so will demand no medical care. She also faces probability, p, also equal to 0.5, that she will contract an illness that requires medical care. Due perhaps to a family history, Elizabeth fears that she will contract Type 1 diabetes—if so, without insulin she will die. In Figure 8.4, panel A, we assume that her demand for insulin is perfectly inelastic, that is, unresponsive to its price—diabetics do not buy more insulin simply because it is cheaper. We saw earlier (ignoring the transaction costs) that Elizabeth would be willing to pay insurance to cover expenditures P_1Q_1, her expenditures if she needs care. An actuarially fair insurance policy would then charge Elizabeth $\frac{1}{2}P_1Q_1$, and she would purchase the policy to insure her against the risk of diabetes.

Consider instead Elizabeth's demand for dermatological care (skin care for conditions such as acne or psoriasis). Elizabeth's demand curve for these elective treatments may very

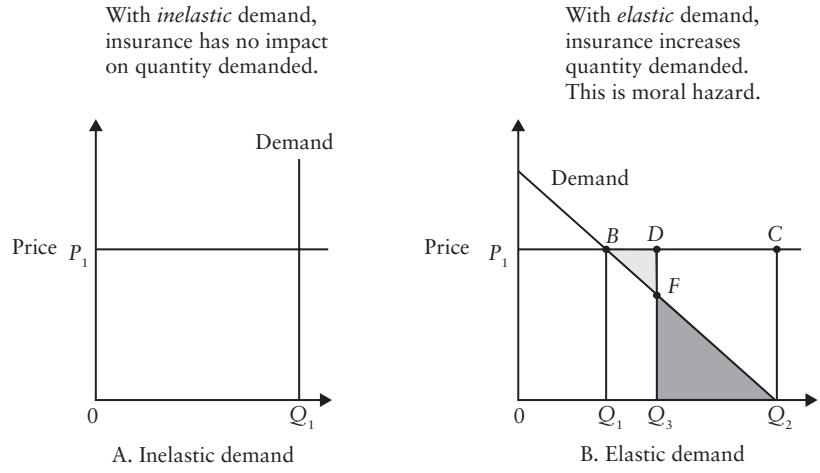


Figure 8.4 Demand for Care and Moral Hazard

well respond to price; that is, the lower the price, the higher the quantity demanded. This is noted in Figure 8.4, panel B. If she purchases insurance that pays her entire loss, this insurance makes treatment (ignoring time costs) free. Because the marginal price to Elizabeth is zero, she would demand Q_2 units of care for a total cost of care of P_1Q_2, shown as rectangle $0P_1CQ_2$, which is obviously larger than rectangle $0P_1BQ_1$. Why only Q_2 units when the care is "free"? Even free care entails time costs of visiting the provider or filling the prescriptions that keep the full price from equaling zero.

The responsiveness of quantity to price leads to one of two possibilities that was not a problem either in the abstract or for a condition like diabetes:

1 If the insurance company charges the premium $\frac{1}{2}P_1Q_1$ (where ½ refers to the probability of illness) for the insurance, the company will lose money. This occurs because the expected payments would be $\frac{1}{2}P_1Q_2$. Amount P_1Q_2 exceeds P_1Q_1 because the induced demand leads Elizabeth to consume more care (Q_2) with insurance than she would have consumed (Q_1) without insurance.
2 If the insurance company charges the appropriate premium, $\frac{1}{2}P_1Q_2$, for the insurance, Elizabeth may not buy insurance. The insurance expense $\frac{1}{2}P_1Q_2$ may exceed the medical expenses that she would have spent on average had she chosen to put away money on her own, or to "self-insure." While Elizabeth may be willing to pay more than $\frac{1}{2}P_1Q_1$ to avoid the risk, she may not be willing to pay as much as $\frac{1}{2}P_1Q_2$.

The rational response to economic incentives brought about by the price elasticity of demand is termed *moral hazard*, the change in consumer behavior in response to a contractual arrangement (here, the decision to insure). In this case, usage of services increases because the pooling of risks decreases the consumer's marginal costs. Failure to protect oneself from disease, because one has health insurance, would be another form of moral hazard (see Box 8.1 for an example). Our analysis gives a simple measure of the economic costs of moral hazard. Netting out the costs of servicing the insurance (which do not reflect increased use of services), moral hazard is the excess of premiums over Elizabeth's expected outlays had she not purchased insurance.

Elizabeth's insurance premium thus has two parts. The first is the premium for protection against risk, assuming that no moral hazard exists. The second is the extra resource cost due to moral hazard. As before, Elizabeth chooses insurance coverage q^* by weighing marginal costs against marginal returns, whereas before the returns were the utility gains when Elizabeth was ill. The twist here is that the costs now have two dimensions—the pure premium and the moral hazard. For some categories of care, the second may be important.

This analysis has helped to predict the types of insurance likely to be provided. It is clear that the optimal level of insurance will likely increase relative to the expected loss as the degree of moral hazard decreases. Suppose we use demand price elasticity to indicate the potential for moral hazard. Theory then suggests:

1 Deeper (more complete) coverage for services with more inelastic demand.
2 Development of insurance first for those services with the most inelastic demand, and only later for those with more elastic demand.

Data on current insurance coverage by area of service support the first hypothesis, and historical data support the second.

BOX 8.1

Another Type of Moral Hazard—Health Insurance and Insecticide-Treated Bed Nets in Ghana

Although health insurance scholars worry about insurance price effects, health insurance may also influence disease prevention efforts. Yilma, van Kempen, and de Hoop (2012) examine the impact of the Ghanaian National Health Insurance Scheme (NHIS) on households' efforts in preventing malaria. The National Health Insurance Act 650 was passed in August 2003 to improve access and quality of basic health care services through a National Health Insurance implemented at the district level. By the end of 2008, every district had enrolled and 61 percent of the total population was covered. The financing of the NHI includes premiums paid by the insured and the NHI fund that comes from taxes on goods, social security contributions, parliament budget allocation and returns from investment.

Sleeping under an insecticide-treated bed net (ITN) is a prominent malaria prevention strategy in sub-Saharan Africa and parts of Asia. Malaria obviously brings forward a utility loss, and possibly death, but people view sleeping under nets as inconvenient. Participants in a community meeting were quoted as saying, "We have mosquito nets but we don't use them. If you are insured it is easier to go to the hospital [in the case of malaria]." Or, "Why would you spend 8 Ghanaian Cedis [currency] on the bed net while you can take 2 Cedis to go to the hospital?"

In mixed statistical analyses the authors found that health insurance negatively impacted bed net ownership, number of members who slept under an ITN, and the number who slept under an ITN they got re-soaked (with insecticides) after they bought it. While the authors do not have firm evidence on whether the incidence of malaria had increased, they have shown that the insurance *for hospital care* reduced levels of user self-protection, unintended consequences from a contractual arrangement.

Effects of Coinsurance and Deductibles

This analysis also provides insight into the impacts of deductible provisions and coinsurance in insurance policies. Returning to Figure 8.4, panel B, suppose that Q_1 reflects \$500 of expenses (rectangle $0P_1BQ_1$) and that Q_2 is three times Q_1 (rectangle $0P_1CQ_2$), which reflects \$1,500 of expenses. If the insurance contains a deductible, Elizabeth will compare the position she would attain if she covered the deductible and received level Q_2 free, with the position she would attain if she paid the market price for all the medical care she consumed.

Assume again that the probability of illness p equals 0.5. Consider first a policy containing a deductible, which requires Elizabeth to pay the risk premium plus the first \$500 of her medical care (expenses indicated by rectangle $0P_1BQ_1$), after which all additional care is free. Elizabeth will buy this policy because it protects her from risk and allows her to purchase Q_2 units of medical care for \$500. Her gain is the triangle under the demand curve, Q_1BQ_2.

Suppose now that the insurance company raises the deductible from $500 to $700. Will Elizabeth continue to buy the insurance? Recall that without insurance, Elizabeth would have purchased amount Q_1 of health services; the $700 deductible yields amount Q_3. When ill, Elizabeth is paying more for the amount $(Q_3 - Q_1)$ of incremental health care than she believes the value of incremental care to be. The incremental costs are rectangle Q_1BDQ_3; the incremental benefits are the area under her demand curve (trapezoid Q_1BFQ_3). The difference is triangle *BDF*, and this represents a welfare loss to Elizabeth. However, after paying the deductible, she can get as much additional health care as she wants at zero cost, and she will buy quantity Q_2. This yields welfare gain triangle Q_3FQ_2 (incremental benefits less zero incremental costs). If Q_3FQ_2 (her welfare gain) is larger than *BDF* (her welfare loss), she buys the insurance, even with the $700 deductible. If *BDF* is larger than Q_3FQ_2, the loss exceeds the gain, and Elizabeth is better off self-insuring and spending P_1Q_1 (in this example, $500) with probability 0.5.

Hence, the deductible has two possible impacts. A relatively small deductible will have no effect on individual usage, here Q_2. A large deductible makes it more likely that individuals will self-insure and consume the amount of care Q_1 they would have purchased with no insurance.

A wide range of coinsurance coverages have developed. Many analysts have considered how to formulate coverages to promote more economically efficient outcomes. We turn next to that analysis.

Health Insurance and the Efficient Allocation of Resources

This section examines the impact of health insurance on health care demand. Economists commonly examine the efficient allocation of resources, which occurs when the incremental cost of bringing the resources to market (marginal cost) equals the valuation in the market to those who buy the resources (marginal benefit). As we first learned in Chapter 4, if the marginal benefit is greater (less) than the marginal cost, one could improve society's welfare by allocating more (fewer) resources to the sector or individual, and less (more) resources to other sectors.

Consider Figure 8.5, which shows the marginal cost of care at P_0 and the demand for care for Elizabeth under alternative conditions of insurance. If Elizabeth is not insured, then the optimal choice of health care is Q_0 units. The price (including travel time, parking, and the cost of bringing the service to market) reflects the cost to society of bringing the entire package to the market. Based on Elizabeth's (and the physician's) preferences, the marginal benefit, as described through the demand curve, equals the marginal cost. In economic terminology, this is an efficient allocation.

The Impact of Coinsurance

What happens when Elizabeth pays only a small fraction of the bill, say, at a 20 percent coinsurance rate? If, for example, P_0 was $50 for an office visit, Elizabeth must now only pay P_1, or $10, so her quantity demanded will increase. This is as if a new demand curve (labeled with 20 percent coinsurance) were generated by rotating the original demand curve outward, and leading to a new equilibrium quantity demanded Q_1. The cost of bringing

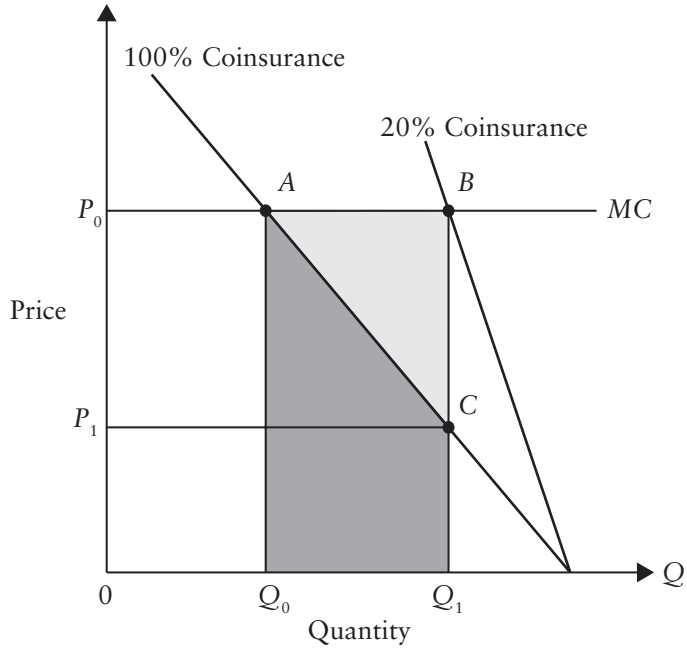


Figure 8.5 Health Care Demand with Insurance

services to market has remained the same, P_0. Services valued at P_0Q_0 are now being provided. The incremental amount spent (incremental cost) is P_0 ($Q_1 - Q_0$), or the rectangle ABQ_1Q_0.

The incremental benefit (to Elizabeth) is the area under her original demand curve, ACQ_1Q_0. The remaining triangle ABC represents the loss in well-being that occurs because Elizabeth is purchasing more health care than is optimal. It is a loss because the incremental resource cost ABQ_1Q_0 exceeds the incremental benefits ACQ_1Q_0 by triangle ABC.

What exactly does this mean? It means that the insurance leads Elizabeth to act as if she was not aware of the true resource costs of the care she consumes. It also means that the insurance essentially subsidizes insured types of care (organized health care settings, prescription drugs) rather than other types of health care (e.g., good nutrition, exercise, over-the-counter drugs, uninsured types of care) that may be just as good, or even better. It also subsidizes insured health care relative to nonhealth goods. The degree of this distortion depends on the exact specification (deductibles, maximum payments, rates of coinsurance) of the policy, but it suggests that insurance can distort the allocation of resources among health care and other goods.

Until recently, many insurance policies had flat rate copayments as low as one or two dollars for all drugs, leading to circumstances under which it could cost more to drive to the pharmacy than to pay for the drugs themselves. Then some insurers instituted two-tiered policies such as 5–10 policies, charging $5 for generic drugs and $10 for brand-name drugs. Box 8.2 examines recent changes in coinsurance rates for prescription drugs with four or five tiers. Tier 4 drugs, in this account, often come with coinsurance rates of 25 percent or higher.

BOX 8.2

Got Insurance? You Still May Pay a Steep Price for Prescriptions

Having health insurance may help pay for drugs, but they may still be very expensive. Writer Julie Appleby recounted the experience of Ms. Sandra Grooms, a general manager at a janitorial supply company in Augusta, GA. The chemotherapy drugs that her oncologist (cancer specialist) wanted to use on her metastatic breast cancer were covered by her health plan, but with one catch: Her share of the cost would be $976 for each 14-day supply of the two pills.

Grooms's response? "I said, 'I can't afford it.'"

Appleby notes that, health plans—even those offered to people with job-based coverage—increasingly require hefty payments by patients like Grooms. Some require coinsurance rates of 20 to 40 percent or more of the total cost of medications deemed to be "specialty drugs." These practices place the drugs in the highest tiers of patient cost sharing. While there may be an out-of-pocket maximum, for many health plans it is often several thousand dollars.

Some patient advocates, writes Appleby, fear that insurers are using high coinsurance rates to skirt the Affordable Care Act's rules requiring them to accept all enrollees, including those with medical conditions. Their logic: while not rejecting anyone, these plans can discourage patients with health problems from enrolling if they set high payments for drugs for specific medical conditions. Insurers often place specialty drugs, which have no standard industry definition but are generally the most expensive products, into the higher tiers. Many patients do not have lower cost alternatives.

Grooms did. Her oncologist selected a different drug—an intravenous medication—for which her cost share is $100 a month, as opposed to $1,952 for the higher-tiered alternative.

Source: Appleby, Julie, "Got Insurance? You Still May Pay a Steep Price for Prescriptions," *Kaiser Health News*, October 13, 2014, http://khn.org/news/got-insurance-you-still-may-pay-a-steep-price-for-prescriptions/, accessed September 12, 2015.

The impact of moral hazard is intensified through interactions between primary and secondary insurance coverages. This type of interaction sometimes describes "Medigap" plans, which provide additional coverage to the elderly above the amount paid by Medicare. Another example involves insured employees who have secondary coverage through their spouse's insurance which may magnify moral hazard problems.

Elizabeth's employer, General National, provides health insurance to all its workers, with policies that pay 60 percent of all medical expenditures. Many of General's workers also receive coverage under their spouse's insurance plans, but General's plan is considered the primary insurer for these dually covered workers. The *secondary* policies cover 60 percent of the expenses left uncovered by General's plan.

Figure 8.6 shows a demand curve for visits for the typical General National family if they had no insurance. The family would spend $600 on 12 visits per year, at a price of $50. If General National is the primary insurer, the out-of-pocket price to its insured will fall by 60 percent to $20 per visit. As drawn, the lower out-of-pocket price to patients increases

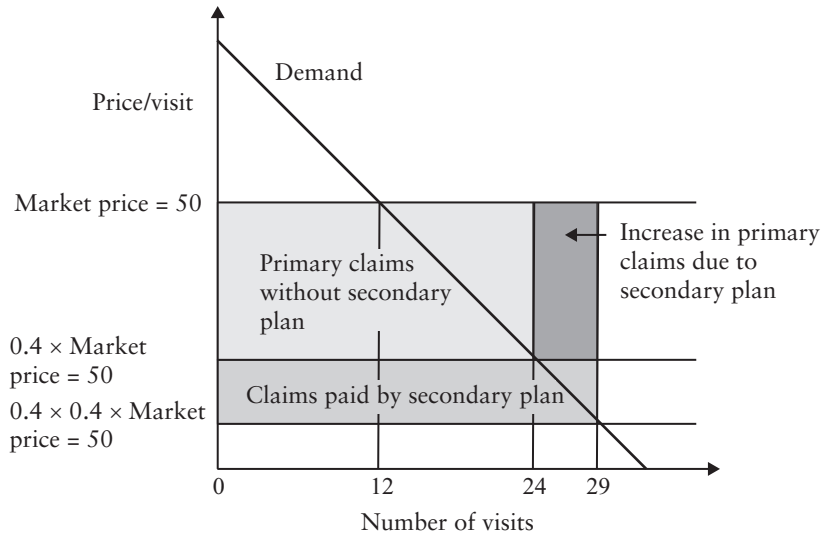


Figure 8.6 Impact of Secondary Insurance on Primary Coverage and Utilization

quantity demanded to 24 visits. General National will pay $720, or 60 percent of the $1,200 total cost; its employees will pay $480.

Consider, however, the impact of secondary insurance. By paying 60 percent of the remainder, the secondary insurers reduce the out-of-pocket cost to the employees by another 60 percent, from $20 per visit to $8 per visit. Not surprisingly, the quantity of visits demanded increases again, this time from 24 to 29 visits. The secondary insurers pay $12 per visit, or $348 for the 29 visits. Moreover, the primary insurer, General National, faces increased claims due to demand induced by the coverage of the secondary insurers. General's liability increases from 60 percent of the original $1,200 in expenditures to 60 percent of $1,450 in expenditures—the higher level resulting from the secondary coverage.

A combination of coverages, while providing additional employee benefits, exacerbates the moral hazard problem brought on in general by health insurance. The inefficiencies and welfare losses that occur when decisions of one firm increase the health care costs facing another pose a difficult problem for policymakers.

Losses may be even more significant in the market context than in the individual context, as described by Figure 8.7. Clearly, as before, more services are used than are optimal. This comprises both a redistribution of resources (from consumers and insurers to providers) and a deadweight loss (referring to a loss that comes from the misallocation of resources between types of goods). At the original price, P_0, and quantity, Q_0, producers were covering the marginal cost to bring the products to market.

The deadweight loss comes from a misallocation of resources among goods (i.e., more health care is provided than should be, according to consumer preferences). Trapezoid Q_0JKQ_1 indicates the incremental benefits induced by the establishment of a coinsurance regime (i.e., the area under the original demand curve).

Similarly, the additional resource costs of bringing the treatment level ($Q_1 - Q_0$) to society is trapezoid Q_0JFQ_1. The deadweight loss from the insurance-induced overproduction of health services is the difference in areas between the two trapezoids, or triangle *FKJ*.

The Demand for Insurance and the Price of Care

Martin Feldstein (1973) was among the first to show that the demand for insurance and the moral hazard brought on by insurance may interact to increase health care prices even more than either one alone. Insurance is related to the expected loss; in health care, this is related in part to the price of care. Increased price of care is related to an increased demand for insurance, as noted in the upward-sloping *I* curve in Figure 8.8.

The second impact is that of insurance on the price of care. More generous insurance and the induced demand in the market due to moral hazard lead consumers to purchase more health care. Line P_1 shows that if the supply curve for health care is horizontal, then increased insurance will not increase the price of care above PC_1. The equilibrium is at point *A*, with health care price PC_1 and insurance quantity Q_1.

If, however, the product supply curve rises, more generous insurance causes market price to increase. We trace this impact as curve P_2. Start at point *A*. The increased product price (the vertical arrow) due to the moral hazard brought on by insurance leads to an increased demand for insurance (the horizontal arrow), which again feeds back on price of care and so on. The moral hazard together with the upward-sloping product supply curve leads to a new equilibrium, *B*, with higher price of care, PC_2, and higher quantity of insurance, Q_2. The combination of factors leads to a higher price of health care and a higher demand for insurance than would have occurred were there no insurance. Many feel that technology-induced price increases along with improved insurance have further increased the price of care.

The Welfare Loss of Excess Health Insurance

From the preceding discussion, one would ask why society would support policies that seem only to result in misallocations of resources. In fact, the foregoing analyses concentrate only on the costs. We emphasize that people willingly buy insurance, taking on additional costs to themselves, to protect against the risk of possibly substantial losses.[3] This protection provides major benefits through the protection against risk; the benefits from protection against risk offset the losses discussed here.

Martin Feldstein (1973) was one of the earliest and most prominent of the researchers attempting to calculate the welfare losses of excess health insurance. He measured the cost of the excess insurance by measuring the demand for health care and the coinsurance rate, and calculating the size of the loss polygons in Figure 8.7.

Measuring the benefits also is straightforward conceptually. One can use Figure 8.1 to measure the horizontal difference between Elizabeth's expected utility and her actual utility. This represents the dollar amount that she would have been willing to pay for insurance over and above the amount that she was charged. Provision of insurance to Elizabeth at the actuarially fair rate provides a utility gain to Elizabeth through the reduction of uncertainty. Adding these gains across individuals provides a measure of net benefits to compare to the costs of the excess insurance.

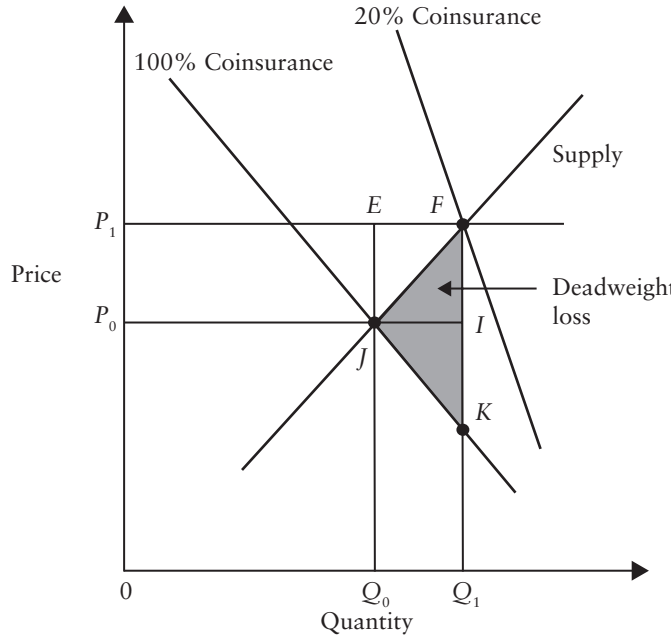


Figure 8.7 The Effect of Insurance Cost Sharing with Upward-Sloping Supply

In plain terms, insurance policies impose increased costs on society because they lead to increased health services expenditures in several ways:

- increased quantity of services purchased due to decreases in out-of-pocket costs for services that are already being purchased;
- increased prices for the services that are already being purchased;
- increased quantities and prices for services that would not be purchased unless they were covered by insurance; or
- increased quality in the services purchased, including expensive, technology-intensive services that might not be purchased unless covered by insurance.

Any procedures that raise the coinsurance rate will tend to reduce the costs of excess insurance but also will reduce the benefits from decreased risk bearing. Feldstein found that the average coinsurance rate was about one-third, or 0.33; that is, people paid $0.33 of every $1 of costs out of their own pockets. Raising the coinsurance to 0.50 or to 0.67 would cut the amount of insurance purchased, reducing the excess insurance, but also increasing the amount of risk borne by the clients.

The welfare gains from changed coinsurance, then, are:

$$\text{Welfare gains} = \text{Change in benefits} - \text{Change in costs} \tag{8.8}$$

Feldstein's analysis considers the welfare gains from increasing the average coinsurance rate from 0.33 to 0.67. He discovers that the costs fall much more than do the benefits as coinsurance rates rise. He estimates the welfare gains to be approximately $27.8 billion per year (in 1984 dollars) under the "most likely" parameter values.

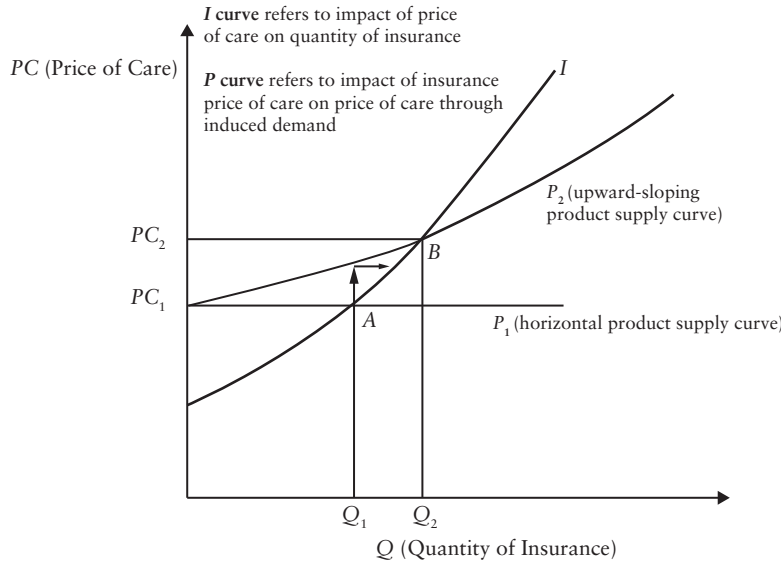


Figure 8.8 The Interaction of Insurance and Price of Care

Feldman and Dowd (1991) updated Feldstein's 1960s estimates with 1980s parameters from the RAND Health Insurance Experiment regarding both price elasticity of the demand curve and attitudes toward risk. They calculate a lower bound for losses of approximately $33 billion per year (in 1984 dollars) and an upper bound as high as $109 billion. For perspective, the upper and lower bounds constituted between 8.9 and 29.1 percent of all 1984 health expenditures.

Manning and Marquis (1996) sought to calculate the coinsurance rate that balances the marginal gain from increased protection against risk against the marginal loss from increased moral hazard, and found a coinsurance rate of about 45 percent to be optimal. Although the impacts of proposed changes depend crucially on the underlying econometric estimates (see Nyman, 1999, for further discussion), the fact that current coinsurance rates have remained far lower than 45 percent suggests a potentially important role for restructuring insurance to reduce excess health care expenditures.

Kowalski (2015) examines the balance between moral hazard and risk protection using the kinds of employer-sponsored insurance policies that are common in the United States. Figure 8.9 shows a health insurance policy where health care is measured in units (or dollars) on the x-axis. Patients pay the full deductible up to level h_1. From h_1 to h_2, they pay at coinsurance rate c (which is less than 1), and past h_2 a "stop-loss" occurs where all of the expenses are covered, so-called catastrophic insurance. Up to h_1 there is no moral hazard, because patients are paying full price. Past h_1, both moral hazard and risk protection occur, with the risk protection being particularly important for very large expenditures.

Using careful theoretical and econometric models, Kowalski finds that the welfare gain from risk protection is "really small at every point in the distribution – on the order of pennies." In contrast the deadweight loss is about 100 times larger. That said, she characterizes both quantities as empirically "very small" relative to the money at stake.

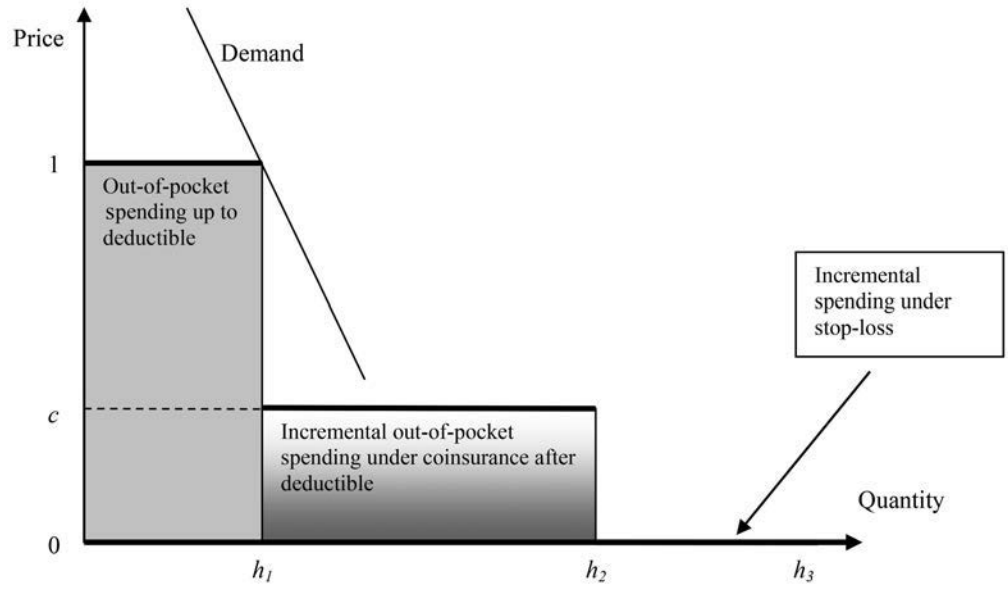


Figure 8.9 Moral Hazard and Risk Protection with Employer-Sponsored Plans

Income Transfer Effects of Insurance

John Nyman has argued (1999) that in addition to the conventional insurance theory, we should also view insurance payoffs as income transfers from those who remain healthy to those who become ill, and that these income transfers generate additional consumption of medical care and potential *increases* in economic well-being. Transfers occur because for most medical procedures, especially expensive procedures for serious illnesses, demand depends on the person's becoming ill.

The conventional analyses following Pauly's 1968 model imply that:

1 there are no income transfer effects due to insurance, and
2 all moral hazard is due to pure price effects.

What healthy consumer, asks Nyman, would purchase a coronary bypass procedure (or bowel resection or organ transplant) just because insurance is available and the price has dropped to zero? Although the prices of such procedures may fall for all who purchase insurance, only those who are ill will respond to the reduction. Because only the ill respond, the price reduction is the vehicle by which income is transferred from the healthy to the ill.

Figure 8.10 presents the corresponding demand curve analysis with the original demand curve labeled D.[4] For the type of medical procedures in question, D represents the behavior of only those who become ill. If the price equals 1, quantity m_u is demanded, but if the price falls to coinsurance rate c, then quantity m_e will be consumed. The demand curve D^i illustrates the effect of the insurance contract on the behavior of the consumer who purchases insurance with a coinsurance rate of c and becomes ill, and consumes an amount of medical care equal to m_i.

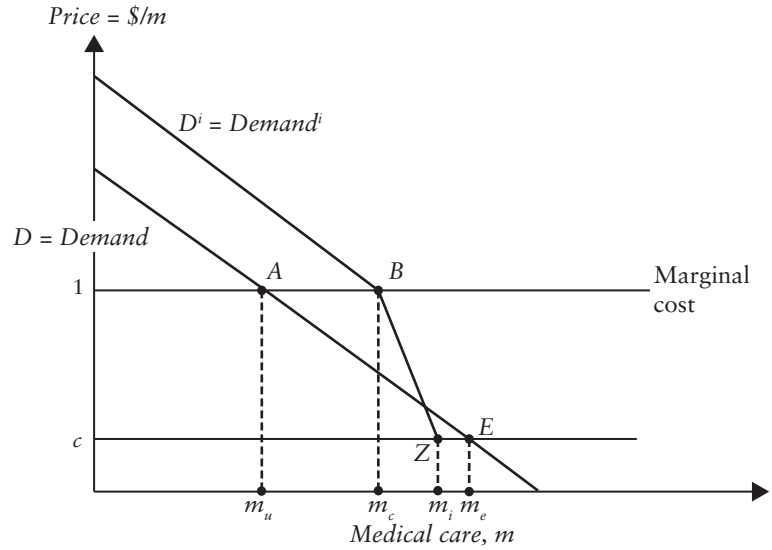


Figure 8.10 Nyman's Decomposition of Moral Hazard Using Demand Curves

The insurance price decrease is the vehicle for transferring income to the consumer who becomes ill. Figure 8.10 shows this income transfer as the portion of D^i that lies above the original price because, if the income were simply transferred, the increase in willingness to pay would shift out demand at any of those prices. For any given probability of illness, the smaller the coinsurance rate that is purchased in the contract, the greater will be the income transfers and the shift in D^i compared to the original demand, representing the conventional response to an exogenous price change.

The portion of D^i that lies below the original price of 1 reflects both behavior and the mechanics of the insurance contract. This portion of D^i is steeper than the original demand because in order to purchase an insurance contract with successively lower coinsurance rates, the consumer must pay successively greater premiums. That is, two changes occur simultaneously as successively lower coinsurance rates are purchased:

- First, the lower coinsurance rates generate a larger transfer of income to the ill consumer, causing the portion of D^i above the original price leading to shift horizontally and to the right, leading to point B.
- Second, the larger premiums associated with lower coinsurance rates generate an ever larger differential between demand curves D and D^i.

For example, purchasing a coinsurance rate of $c < 1$ requires a premium payment that causes a demand differential equal to the horizontal difference between points E and Z in Figure 8.10 due to the assumed responsiveness to income. The purchase of a lower coinsurance rate would produce an even larger horizontal difference.

Figure 8.11 illustrates the gain from insurance for the ill consumer who purchases an insurance policy with coinsurance rate 0. The income transfer increases willingness to pay for medical care, shifting out the portion of demand that is above the existing market price of 1. This results in an increase in the consumer surplus of area $FBAG$. However, we must

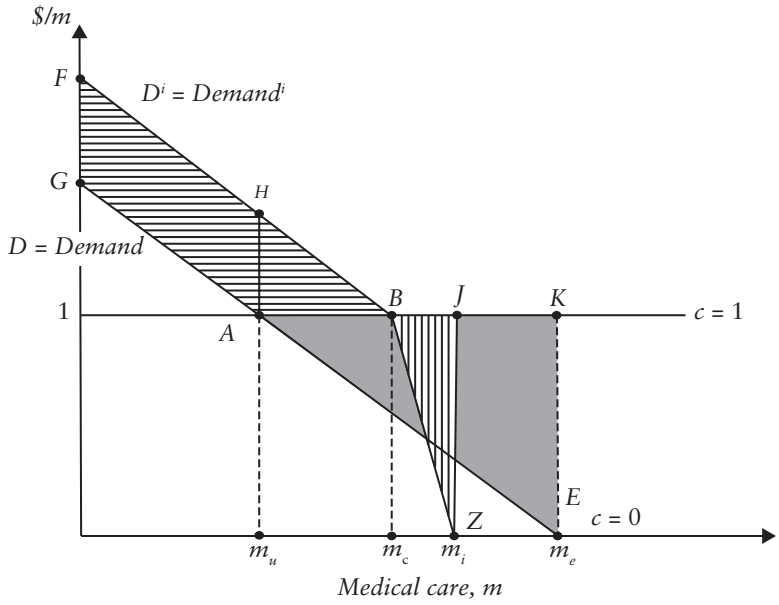


Figure 8.11 The Net Welfare Gain of the Price Distortion under Income Transfer Effects

subtract welfare loss *BJZ* generated by using a price reduction to transfer this income. Calculate the net welfare gain by subtracting area *BJZ* from area *FBAG*. In comparison, under conventional theory insurance only produces a welfare loss, which would be represented by area *AKE* (shaded in gray).

Here is a numerical example. Suppose that Elizabeth receives a diagnosis of breast cancer at an annual screen. Without insurance, she would purchase a mastectomy for $30,000 to rid her body of the cancer. In this example Elizabeth has purchased insurance for $6,000 that pays for all her care (zero coinsurance rate means that $c = 0$) if ill. With insurance, Elizabeth purchases (and insurance pays for) the $30,000 mastectomy, a $20,000 breast reconstruction procedure to correct the disfigurement caused by the mastectomy, and two extra days in the hospital to recover, which costs $6,000. Total spending with insurance is $56,000 and total spending without insurance is $30,000, so it appears that the price distortion has caused $26,000 in moral hazard spending.

Is this spending increase truly inefficient? We must determine what Elizabeth would have done if her insurer had instead paid off the contract with a cashier's check for $56,000 upon diagnosis. After accounting for her $6,000 premium payment, the payoff represents ($56,000 – $6,000) or $50,000 in income transfers that she could spend on anything of her choosing. With her original resources plus the additional $50,000, assume that Elizabeth would purchase the mastectomy and the breast reconstruction, but not the extra days in the hospital. In Figure 8.11, the mastectomy would be represented by m_u, the breast reconstruction by $(m_c - m_u)$, and the two extra days in the hospital by $(m_i - m_c)$. This implies that the $20,000 spent on the breast reconstruction is efficient and welfare-increasing, but the $6,000 spent on the two extra hospital days (induced by the zero copayment) is inefficient and welfare-decreasing, consistent with the conventional theory.

Nyman's work provides an important extension to the theory of health insurance. The income effects that he identifies are justifiable additions to economic welfare, and we should net them out against potential excess costs brought on by moral hazard.

Conclusions

This chapter has concentrated on the unique role of insurance in the health care economy. No other good in consumers' budgets is so explicitly tied to the arrangements for insurance. Health insurance arrangements affect not only expenditures for serious illnesses and injuries, but also plans for more routine expenditures, such as children's well-care visits (for infants and toddlers), and eye and dental care.

We have characterized risk and have shown why individuals will pay to insure against it. Under most insurance arrangements, the resulting coverage leads to the purchase of more or different services than might otherwise have initially been bought by consumers and/or their health care providers. Health care policy experts focus on how to structure insurance policies in order to reduce purchases and minimize insurance costs without compromising the health of the insured.

Summary

1 Many illnesses occur rarely and seemingly at random, but when they do, they entail substantial costs.
2 Insurance reduces variability of people's assets by creating large pools of customers and operating according to the law of large numbers. Although outlays for a health event may vary significantly for any given unit in the pool, average outlays for the group are fairly predictable. If they are predictable, they can be insured.
3 One should distinguish between insurance, as is provided through the pooling of risk, and government programs, such as Social Security, Medicare, and Medicaid, which also redistribute wealth.
4 Insurance can be sold only in circumstances with diminishing marginal utility of wealth or income (i.e., when the consumer is risk-averse). With constant marginal utility, actuarially fair premiums would leave consumers no better off than if they were uninsured.
5 Expected utility is an average measure; the individual either wins or loses the bet.
6 If insurance companies charge more than the actuarially fair premium, people will have less expected wealth through insuring than through not insuring. Even though people will be less wealthy by purchasing insurance, the increased well-being comes from the elimination of risk.
7 In theory, the optimal amount of insurance in the absence of loading costs leads to full insurance against the expected loss. With loading costs, the optimal coverage is less than the expected loss.
8 Moral hazard refers to the increased usage of services when the pooling of risks leads to decreased marginal price for the services. This suggests:
   - more complete coverage for price inelastic services, and
   - earlier development of insurance for services that are most inelastic.
9 Insurance policies increase costs to society because they increase expenditures on health services. They provide increased benefits through the reduction of risks.

10 Some analysts have calculated losses due to excess insurance as between 8.9 and 29.1 percent of all health expenditures. This suggests the importance of restructuring insurance to reduce excess health care expenditures.
11 Nyman shows that under many circumstances insurance payoffs represent income transfers from those who remain healthy to those who become ill. In these cases, the income transfers generate welfare-increasing additional consumption of medical care.

Discussion Questions

1 Discuss the difference between cardinal and ordinal utility. Why is cardinal utility necessary for the analysis of risk and insurance?
2 What does the term *moral hazard* mean? Give examples.
3 The deductible feature of an insurance policy can affect the impact of moral hazard. Explain this in the context either of probability of treatment and/or amount of treatment demanded.
4 Describe the benefits to society from purchasing insurance. Describe the costs. Define and discuss the welfare gains from changes in insurance coverage.
5 If only risk-averse people will buy health insurance, why do many people who buy health insurance also buy lottery tickets (an activity more consistent with risk-taking, especially since most lotteries are not actuarially fair)? Speculate on the differences and similarities.
6 The game show *Deal or No Deal*, popular throughout the world, provides many elements of risk and expected value. Discuss the ways that the decision as to whether to "take the money" or to continue involves evaluation of risk and expected value. In the episode www.youtube.com/watch?v=H9CQscwXBt0, is the decision making rational or not?
7 Some brokers (called viatical brokers) offer cash settlements in advance to people with terminal diseases who have life insurance, paying them in advance of their death. Is this practice ethical? Is it ethical for the brokers to offer advance settlements to elderly people simply because they may die soon?
8 Because health insurance tends inevitably to cause moral hazard, will the population necessarily be overinsured (in the sense that a reduction in insurance would improve welfare)? Are there beneficial factors that balance against the costs of welfare loss?
9 From Nyman's arguments, do all increased expenditures become welfare enhancing? Give examples of some that enhance welfare. Give examples of others that do not.

Exercises

1 Suppose that Nathan's employer provides a health insurance policy that pays 80 percent of $1 over the first $100 spent. If Nathan incurs $1,000 in expenses, how much will he pay out-of-pocket? What percentage of his expenses will this be?
2 Suppose that rather than flipping a coin, one rolls a die. If the value is 1, 2, 3, or 4, the player wins $1. If it is 5 or 6, the player loses $1. Calculate the expected return.
3 A standard roulette wheel has an array of numbered compartments referred to as "pockets." The pockets are red, black, or green. The numbers 1 through 36 are evenly split between red and black, while 0 and 00 are green pockets. For every $1 that one bets on red, one wins $1 if the roulette ball lands on red and loses if it lands otherwise. Similarly for black. What is the expected return to a red or a black bet? Why?

4 (a) Draw a utility of wealth curve similar to Figure 8.1 for consumers who are not risk-averse. How would its shape differ from Figure 8.1?
   (b) Draw a utility of wealth curve similar to Figure 8.1 for "risk-lovers." How would its shape differ from Figure 8.1?
5 (a) Show the gains from insurance, if any, in Exercise 4a.
   (b) (Difficult) Show the cost of insurance in Exercise 4b.
6 We have discussed the role of utility functions in the purchase of insurance.
   (a) Suppose Edward's utility function can be written as:

   $U = 20Y$ where U is utility and Y is income per month.

   What is his marginal utility if income is \$1,000 per month? \$2,000 per month? Is Edward likely to insure against loss of income? Why?
   (b) Suppose instead that Edgar's utility function can be written as $U = 200Y^{0.5}$. What is his marginal utility if income is \$1,000 per month? \$2,000 per month? Is Edgar likely to buy insurance against loss of income? Why?
   (c) Suppose that Edmund's utility function can be written as $U = 0.5Y^2$. What is his marginal utility if income is \$1,000 per month? 2,000 per month? Is Edmund likely to buy insurance against loss of income? Why?
7 Suppose, if ill, that Fred's demand for health services is summarized by the demand curve $Q = 50 - 2P$, where P is the price of services. How many services does he buy at a price of \$20? Suppose that Fred's probability of illness is 0.25. What is the actuarially fair price of health insurance for Fred with a zero coinsurance rate?
8 In Exercise 7, if the insurance company pays Fred's entire loss, what will Fred's expenses be? How much will the company pay? Will it continue to offer him insurance at the actuarially fair rate? Why?
9 Suppose that the market demand for medical care is summarized by the demand function:

   $Q_d = 100 - 2p$

   and the market supply is summarized by the supply function:

   $Q_s = 20 + 2p$

   (a) Calculate the equilibrium quantity and price, assuming no health insurance is available.
   (b) Suppose that health insurance is made available that provides for a 20 percent coinsurance rate. Calculate the new equilibrium price and quantity. (Hint: How does the demand curve shift?)
   (c) Calculate the deadweight loss due to this insurance.
10 Suppose that the market demand for medical care is summarized by the demand function:

   $Q_d = 200 - 2p$

   and the market supply is summarized by the supply function:

   $Q_s = 20 + 2p$

   (a) Calculate the equilibrium quantity and price, assuming no health insurance is available.
   (b) Suppose that health insurance is made available that provides for a 10 percent coinsurance rate. Calculate the new equilibrium price and quantity.
   (c) Calculate the deadweight loss due to this insurance.
   (d) Compare your answers in this problem to problem 9 (a) – (c).

11 Suppose, in Exercise 9, that the coinsurance rate was raised to 50 percent.
   (a) Calculate the new equilibrium price and quantity. (Hint: How does the demand curve shift?)
   (b) Calculate the deadweight loss due to this insurance.
   (c) How does your answer compare to the deadweight loss in Exercise 9?
12 Consider the discussion in the text about Elizabeth's breast cancer treatment. Using Figure 8.11, calculate the net welfare benefits if m_u = 20,000, m_c = 40,000, and m_i = 44,000. To aid in the calculations, assume that point *G* has a value of 2 and point *F* has a value of 3.

Notes

1 Wealth refers to the sum (or stock) of the consumer's assets in money terms. It is related to income, which is the flow of funds in any given period in that increased income allows people to buy more assets. We may refer to one or the other for some discussions, but the substance of the insurance analysis refers to both.
2 Insurers do care whether they are getting nonrepresentative slices of the risk distribution (suppose, for example, the same member files claims year after year, hardly a random event). Analysts refer to this as *adverse selection*, and it can lead to financial losses for the insurer. We treat this information problem in detail in Chapter 10.
3 Students might ask about people whose employers pay the entire insurance bill. Most economists believe that employees choose the insurance in lieu of a compensating take-home wage, thus paying for insurance themselves. We discuss this in detail in Chapter 11.
4 Nyman and Maude-Griffin (2001) provide the mathematical foundation for this analysis.

Chapter 9

Consumer Choice and Demand

In this chapter

- Applying the Standard Budget Constraint Model
- Two Additional Demand Shifters—Time and Coinsurance
- Issues in Measuring Health Care Demand
- Empirical Measurements of Demand Elasticities
- Impacts of Insurance on Aggregate Expenditures
- Other Variables Affecting Demand
- Conclusions

We have described health capital investment as a choice made by the consumer who values health but who also values the home good, which involves all the goods made possible by using income and time. The consumer produces these ultimate goods, health capital and the home good, by allocating a portion of time to each, as well as a portion of income to those market goods, such as medical care, M, and various other goods, OG, that are used in the production process.

Figure 9.1 helps illustrate this transformation from one model (see Chapter 7) to the other—the conventional analysis of choice over marketable goods. The production possibilities frontier in the figure illustrates the consumer's trade-off between health investment and the home good. Choosing these optimal quantities, I^* and B^*, the consumer also implicitly chooses an allocation of time available for these production tasks, to work that provides income, and to leisure.

Indifference curves U^* and U^{**} provide insights into how different people may choose between the present and the future through a concept known as the *rate of time preference*. Curve U^* represents an indifference curve for Tom. Tom has a high rate of time preference and places a large value on current consumption relative to *future* consumption. The latter is adversely affected by the higher mortality risks and other consequences (e.g., lower future earnings) of poor health. As such, he will choose high present consumption B^* and relatively low health investment level I^*.

His brother Jerry, in contrast, has a low rate of time preference, as noted by curve U^{**}. He places a low value on current consumption relative to future consumption and is more willing than Tom to invest in health. He chose levels B^{**} and I^{**}.

We now wish to examine how income will be used to buy those market goods, such as medical care, that will in turn help the consumer produce health investment and the home good. The consumer faces a trade-off in the marketplace between the use of income to purchase medical care and the use of the money to purchase other goods. This trade-off is the budget constraint of standard consumer theory and is the focus of this chapter.

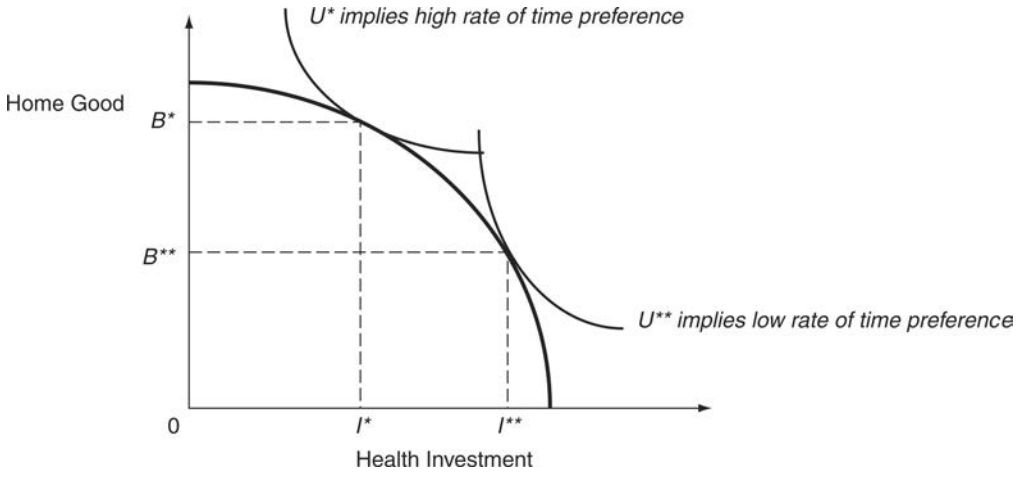


Figure 9.1 Demand for Health Capital Determines the Optimal Amounts of the Home Goods and Health Capital Investment

Applying the Standard Budget Constraint Model

As with the demand for health capital model, standard indifference curve analysis of consumer choice under a budget constraint describes the consumer with fairly strong assumptions. We assume that the consumer is rational and perfectly informed, that there is no uncertainty about the future, and that important decisions are made as if the future were known with certainty. Although we will later relax some of these assumptions, this model produces many reliable predictions on consumer behavior related to health.

Many might object to this approach at the start, believing that we have few choices when we need health care, at least for the urgently ill. If you are lying on the pavement and the ambulance arrives, do you ask for a list of prices and providers?

Nevertheless, we can defend a theory of rational choice over health care and other goods on several grounds. First, many health care options leave room for some thoughtful consideration or at least some planning. (Box 9.1 describes some surprising effects associated with increased patient participation in medical decisions.) Second, the physician serves as an agent for patient-consumers and can make rational choices on their behalf even in urgent situations. Finally, the ultimate test of any theory is whether it predicts well, and we will show empirically that people, as consumers of health care, do respond to economic incentives.

In economic theory, the logic of consumer choice is straightforward. It indicates that consumers can choose any affordable combination or bundle of goods, and from among these affordable bundles, they will choose the most preferred. The depiction of this choice requires two elements:

- The consumer's preferences—described by a set of indifference curves.
- The consumer's budget constraint—described by the straight budget line.

To make the graphical depiction possible, we abstract from the many goods available in the real world and assume instead that only two goods are available. The results for this two-good world generally hold when the model is extended to many goods. Let one of these two goods represent a composite of other goods, and call this good Other Goods, or *OG*. Assume that the health care good is physician office visits consumed during a year, or VISITS. The consumer's name is Ellen Anderson.

BOX 9.1

What Happens to Costs When Patients Participate in Medical Decision Making?

For economists, the efficiency of markets rests on the premise that production responds to consumer preferences. Clinicians and policymakers widely believed that greater patient involvement in their treatment can reduce costs while also improving patient outcomes. Problems may thus arise when the provider, acting as an agent for the patient, helps to determine the amount of care. One type of problem (see Chapter 15), known as supplier-induced demand, occurs when physicians act in part in

their own financial interests rather than their patients' interests. Another, and a seriously understudied problem, arises when communication and other barriers prevent physicians from taking patient preferences into account (Chandra, Cutler, and Song, 2012). What happens to resource use when patients become more involved in their treatment decisions?

Tak, Ruhnke, and Meltzer (2013) surveyed nearly 22,000 patients admitted to a major hospital with questions about their preferences for involvement in medical decisions. The authors used length-of-stay and total hospitalization costs as dependent variables in analyses that controlled for various patient characteristics and self-assessed health status. Among the many results, the investigators found that those who did not agree with the statement "I prefer to leave decisions about my medical care up to my doctor" had about 5 percent higher lengths-of-stay and hospital costs than those who definitely agreed with the statement.

Policies to promote higher levels of patient engagement and shared decision making may thus increase rather than reduce costs. It is important to recognize that the study did not address patient health outcomes or satisfaction with their care.

The Consumer's Equilibrium

Figure 9.2 depicts these elements of the choice problem. The indifference curves labeled U_1, U_2, and U_3 represent some of Ellen's indifference curves (not all are depicted), and together the indifference curves describe her preferences. The indifference curve U_1, for example, represents all points—that is, bundles of OG and VISITS—that provide her with utility level U_1. Utility is an index of preferences that can most easily be understood as a measure of satisfaction. Because OG and VISITS, V, are both goods to the consumer, it follows that indifference curve U_2 is preferred to U_1, and so on; that is, "higher" indifference curves are preferred.

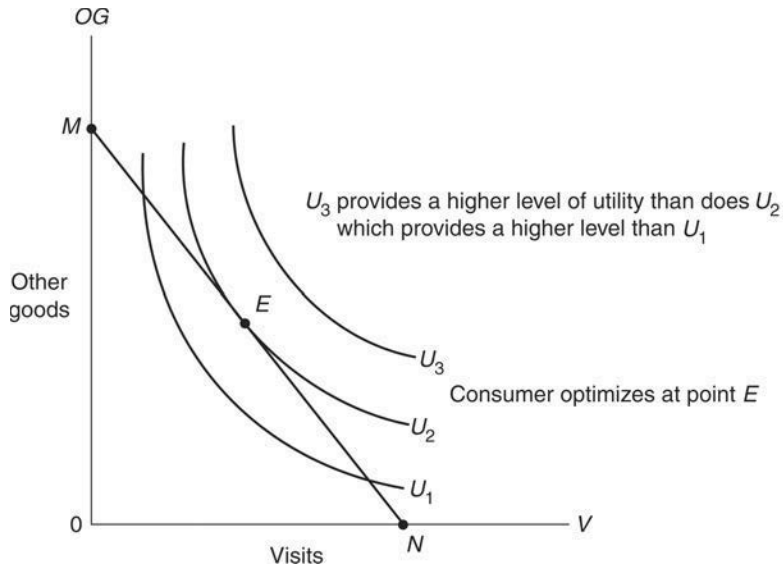


Figure 9.2 Consumer Equilibrium Analysis

Let Ellen's budget be Y dollars for the period. If the price of other goods, OG, is P_{OG} and the price of VISITS is P_V, then the sum of her expenditures, $P_{OG} \times OG$ plus $P_V \times$ VISITS, cannot exceed her income, Y. To spend all her income means to be on the budget line, which is given by the equation:

$$Y = P_{OG} \times OG + P_V \times \text{VISITS}$$

Point M represents the amount of other goods consumed if no visits occur. Point N represents the amount of visits if no other goods are purchased. The budget line is shown as line MN in Figure 9.2, and its slope will be given by $-P_V/P_{OG}$, which is the negative of the ratio of prices.

The consumer equilibrium is point E in Figure 9.2, a point of tangency between the highest indifference curve attainable, U_2, and the budget line. In contrast, all points on indifference curve U_3 are unattainable, and points on U_1 are not chosen because the consumer can afford points she prefers to these. The equilibrium point E is a point of tangency, meaning that the slope of the indifference curve equals the slope of the budget line at this point.

The slope of the indifference curve is called the marginal rate of substitution (MRS). It tells the rate at which Ellen is willing to trade other goods for physician visits. Recall that the slope of the budget line is the negative of the ratio of prices. This is the rate at which she is able to trade other goods for physician visits at current market prices. An equilibrium is reached only if the rate at which she is willing to trade the two goods, the MRS, is equal to the rate at which she is able to trade the two goods, $-P_V/P_{OG}$. This will have the result that in equilibrium, a dollar's worth of OG will yield the same extra (marginal) utility as a dollar's worth of VISITS.

Demand Shifters

Ellen's response to price changes can be determined by examining the new equilibria that would occur as the price of V varies. Figure 9.3 shows the effects of changes in prices at initial income Y, dropping from the highest price, P_V^1, to a lower price, P_V^2, to the lowest price, P_V^3.

At the highest price, P_V^1, income Y buys V_1 visits at equilibrium point E_1. At the lower price, P_V^2, Ellen chooses equilibrium point E_2 (with V_2 visits), and at the lowest price, P_V^3, Ellen chooses equilibrium point E_3 (with V_3 visits). The number of visits, V, increases because visits have become less expensive relative to other goods.

Figure 9.4 plots a demand curve relating the price of visits to equilibrium quantity demanded. The data come from Figure 9.3. Point E_1 from Figure 9.3 corresponds to point A in Figure 9.4, and similarly points E_2 and E_3 correspond to points B and C. The demand curve summarizes Ellen's response to price changes, holding income and preferences constant.

We use price elasticity to measure the responsiveness of the consumer's demand to changes in price. Price elasticity, E_p, is the ratio of the percentage change in quantity demanded to the percentage change in price. Algebraically, it is:

$$\text{Elasticity} = E_p = \frac{(\Delta Q / Q)}{(\Delta P / P)} \tag{9.1}$$

Here, the numerator is the change in quantity divided by the initial quantity level; that is, the numerator is the percentage change in quantity. Because the demand curve slopes downward, the percentage change in quantity (the numerator) will always be negative in

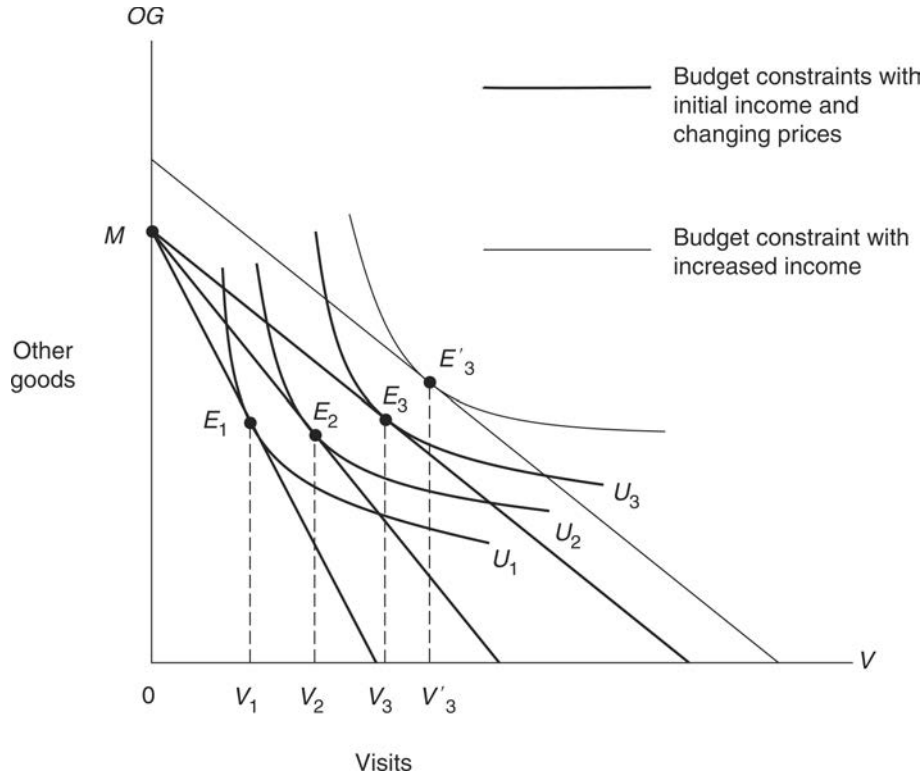


Figure 9.3 Change in Number of Visits as Visit Price and/or Income Changes

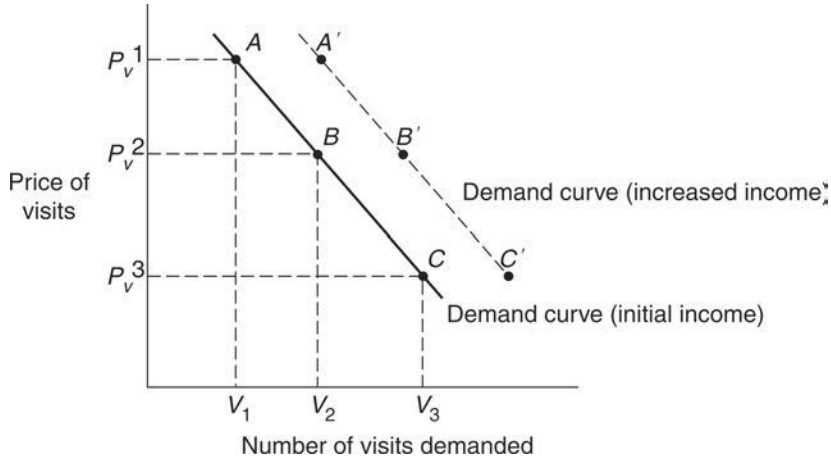


Figure 9.4 Demand Curve Derived from Figure 9.3

response to an increase in price. Likewise, the denominator is the percentage change in price. The higher the elasticity in absolute value (the farther away from 0), the more responsive the consumer is to price. Note that we can write equation (9.1), the *price elasticity of demand*, as:

$$\text{Elasticity} = E_p = \left(\frac{\Delta Q}{\Delta P}\right)\left(\frac{P}{Q}\right)$$

A similar analysis develops the consumer's response to changes in income. Returning to Figure 9.3, recall that point E_3 is determined by income, preferences, and price, P_V^3. Suppose that Ellen's income now increases. Because the relative prices do not change with the income increase, the slope of the budget line does not change, but Ellen can now buy more of both visits and other goods. Her new equilibrium point is E_3'. This translates in Figure 9.4 to point C'. We can similarly draw points A' and B' on Figure 9.4 to indicate the impacts of an income change and prices, P_V^1 (new point A') and P_V^2 (new point B').

The responsiveness of demand to changes in income is measured by the income elasticity. Income elasticity, E_Y, is the percentage change in quantity demanded divided by the percentage change in income:

$$\text{Income elasticity} = E_Y = (\Delta Q / Q)/(\Delta Y / Y), \text{ or}$$

$$E_Y = \frac{\Delta Q}{\Delta Y}\left(\frac{Y}{Q}\right) \tag{9.2}$$

Finally, although two-dimensional indifference curves are not well suited to the handling of larger numbers of substitute and complement goods, the effects of changes in the prices of other goods can be analyzed. One would expect that increases in the prices of substitutes to physician visits (hospital outpatient services, visits to other providers) would increase the demand for office visits. In other words, an increase in the price of a substitute will shift the demand curve to the right in Figure 9.4. Increases in the prices of complements (diagnostic services) would reduce demand for office visits.

Health Status and Demand

Figure 9.5 illustrates how to handle differences in patient health status. Suppose that Ellen is viewed in two different time periods in which her situation is the same in all economic respects except her health status. In Period 1 (equilibrium point E), she is fairly healthy. In Period 2, her overall health status is lower because she is ill. The change in health status will affect Ellen's preferences (often referred to as tastes) for VISITS and OG as reflected by different sets of indifference curves and changed levels of physician care, here point E'.

Note that although Ellen consumes more visits in Figure 9.5 when she is ill, in both cases she has the same MRS of visits for other goods at the equilibrium. Indeed, as long as the prices of the two goods do not change, all consumers in equilibrium will adjust to the prices until all consumers have the same MRS.

The analysis thus far suggests that price, income level, tastes, health status, and other circumstances influence the consumption of physician services. However, we must also consider the roles of insurance and of time.

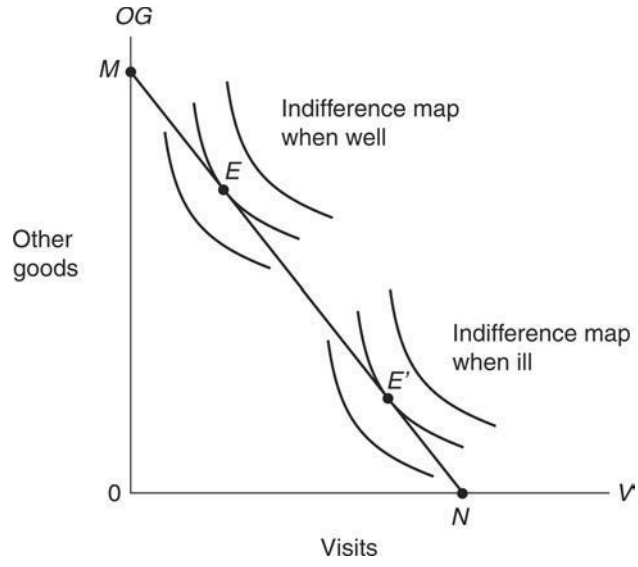


Figure 9.5 Changed Preferences Due to Illness

Two Additional Demand Shifters—Time and Coinsurance

Two demand-shifting variables "look" like changes in the price: time price and coinsurance. First, Ellen's time price, the value of the time she must give up for a physician visit, can represent a significant portion of her full price. Second, insurance reduces Ellen's effective price, the price paid "out-of-pocket," below the market price.

The Role of Time

Recall from Chapter 7 that time is an important element in the demand for health. The consumption of health care services requires considerable time for some services and procedures. Economic observation suggests that people value their time. Many turn down additional work, even at increased wages, such as "time and a half" overtime. Still others decline to drive across town to save $5 or $10 on an item, even though the cost of driving is far less than the $5 or $10. These choices probably occurred because the additional time spent wasn't worth it to the consumer.

Given the opportunity cost of time, a focus on the money costs of health care ignores a substantial portion of the economic costs. The discrepancy between the total economic costs and the money costs will be especially large for low-priced services, for services where patient copayments are small, and for patients with high opportunity costs of time.

As an example of time cost effects, suppose that Ellen must go to the doctor for a 10-minute visit. It will take her 15 minutes to travel each way (30 minutes in all), 20 minutes to wait in the office, and 10 minutes with the doctor. Suppose further that the money

cost of the visit is $25, and that she values her time at $10 per hour. Traveling and parking cost $5 total. The full price of each visit is then $40:

- One hour of time valued at $10.
- One visit priced at $25.
- Travel and parking costs at $5.

Figure 9.6 illustrates that Ellen demands six visits when her full price is $40. A money price increase of $5 causes the new full price to be $45, at which she demands five visits. Restating the price elasticity formula from equation (9.1), we find that the elasticity with respect to the full price is

$$E_p = \frac{(\Delta Q / Q)}{(\Delta P / P)} = \frac{(-1 / 5.5)}{(+5 / 42.5)} = -1.545$$

As appropriate, we use "arc elasticity" to evaluate the price at the midpoint (42.5) between the beginning (40) and the ending (45) price, and similarly for quantity.

Here, P represents the full price; that is, $P = P_M + P_T$. The full price is the sum of money price and time price. In contrast, the money price elasticity in this case is:

$$E_{pM} = \frac{(\Delta Q / Q)}{(\Delta P / P)} = \frac{(-1 / 5.5)}{(+5 / 27.5)} = -1.000$$

In general, the money price elasticity is smaller than the full price elasticity by the same proportion as the money price is smaller than the full price. To make sense of this, try comparing the ratio:

$$P_M \div (P_M + P_T) = 27.5 \div 42.5 = 0.647$$

to the ratio:

$$E_{pM} \div E_P = -1.000 \div (-1.545) = 0.647$$

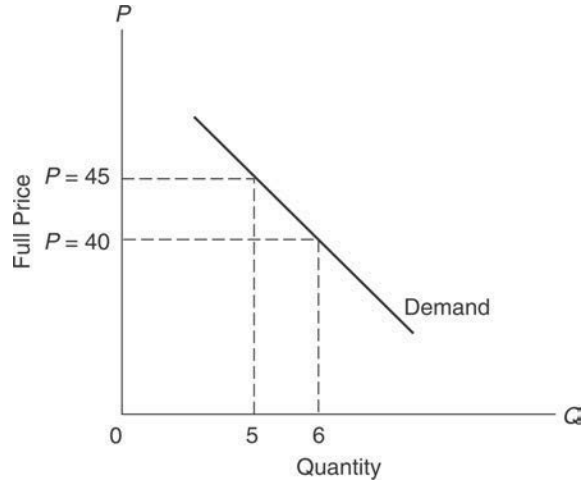


Figure 9.6 Demand and Time Price for Physician Visits

How might this apply? Assuming that the poor have a lower opportunity cost of time than the well-to-do, one would predict that they would more likely tolerate or endure long waiting times in clinics or physician offices. At the same time, even those poor whose physician fees are subsidized (e.g., by Medicaid) must pay their time price. Wishing to increase physician visits among the poor, we might choose to reduce the time price by building nearby clinics and expanding outreach programs, a strategy that has been developed in many localities.

In practice, does time price affect demand? In a pioneering work on time price, Acton (1975, 1976) examines the effects of travel times, waiting times, and other variables on quantity demanded of outpatient visits and physician care. Table 9.1 reports his elasticity estimates. For example, outpatient care, T_{Out}, is the own-time price, and T_{Phys} refers to the other good (cross-time price). The own-time price elasticity for outpatient visits, –0.958, nearly reaches unity, and the own-time price elasticity for physician visits, though small, also shows the importance of time. The positive cross-elasticities suggest that outpatient and physician visits are substitutes.

Subsequent work usually supports an important role for time. Coffey (1983) finds time price also relevant to the decision to seek care initially, as well as the quantity consumed, though her estimates are small (with the exception of public provider care). Mueller and Monheit (1988) find time-price elasticities for dental care to significantly affect the quantity of dental care consumed.

Data from other countries support the analysis. The National Health Service in the United Kingdom, which has eliminated most price constraints on the use of health care resources, finds the remaining waiting time price to be an important rationing factor (Blundell and Windmeijer, 2000). Varkevisser and colleagues (2010) examine non-emergency outpatient visits for neurosurgery in Dutch hospitals (where there are no patient copayments). They found that time elasticities across hospitals were consistently high though they varied widely (–1.4 to –2.6). Does time price affect health care demand? Yes. It makes sense in theory, and it matters in practice.

The Role of Coinsurance

Building on our study of insurance in Chapter 8, we see coinsurance as a demand shifter that works by modifying the effective price. When a third party, such as an insurance company, pays a portion of the hospital bill, the remaining portion paid by the consumer is called the coinsurance rate r. Thus, more insurance means a lower r.

EFFECTS OF REDUCED COINSURANCE ON THE INDIVIDUAL CONSUMER Suppose Ellen has no health insurance and pays all her health care bills out-of-pocket. Figure 9.7 shows Ellen's health care demand as D_0. Because she is uninsured, the market price also is always Ellen's effective (out-of-pocket) price. At a market price of P_1, her quantity demanded is Q_1. A simple thought experiment reveals the issue. Suppose Ellen is given a health insurance policy at no charge (that is, with no impact on the rest of her disposable income) that pays 50 percent of each of her bills, giving her a coinsurance rate of $r = 0.5$. The market price, P_1, is no longer the effective price; the effective price becomes $0.5 \times P_1 = P_1'$. Using her demand curve, D_0, as our guide, we see her now demanding Q_1'. This develops a first principle—her quantity demanded under coinsurance can be found along the out-of-pocket demand curve, provided we identify and apply the effective price.

It is more useful to identify her demand curve with respect to the market price. Ellen acted as though her health care demand had shifted, and this "rotating shift" can be shown to be equivalent to the previous analysis. In Figure 9.7, let us reverse the question and ask: If Ellen demands Q_1' when she has 50 percent insurance coverage and faces an effective price of P_1',

Table 9.1 Acton's Time Valuation Equations

Dependent Variable	*Outpatient Visits*	*Physician Visits*
Elasticity with respect to T_{Out}	–0.958	0.640
Elasticity with respect to T_{Phys}	0.332	–0.252

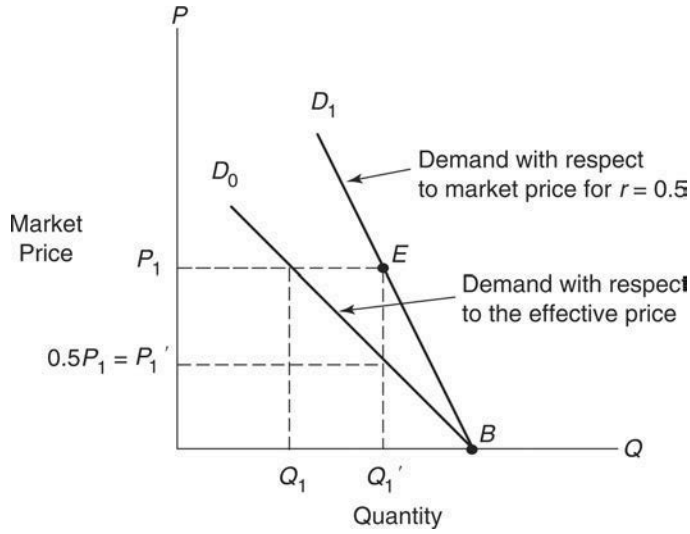


Figure 9.7 The Effect of a Coinsurance Rate on Health Care Demand

then what market price is she apparently willing to pay (part paid by her out-of-pocket and part paid by her through her insurance company)? The answer is P_1, and the resulting combination of quantity demanded and market price is point E in Figure 9.7. By plotting such points (not forgetting the case where $P = P' = 0$), we trace the demand curve with respect to the market price, D_1. Ellen, by receiving the insurance, will act just like an uninsured Ellen whose health care demand curve rotated to the right, hinged at point B.

The exercise makes two theoretical facts clearer: Insurance will increase Ellen's demand for health care, and insurance will make Ellen's demand for health care less elastic. Suppose Ellen's coinsurance rate were zero, meaning she pays nothing for health care. Would her demand be even less elastic? Most health economists would predict that her demand curve would become vertical, hence perfectly inelastic. Since she pays nothing, her demand is totally unresponsive to money price.

MARKET EFFECTS The effect of a reduced coinsurance alone, for Ellen, is an increase in the quantity demanded. Ellen does not demand enough care to influence market prices, and since individual consumers are price takers, their individual actions have no effect on the price. They essentially face a horizontal supply curve. Suppose, however, that the coinsurance rate changes for many consumers in the market. For the market as a whole, the relevant

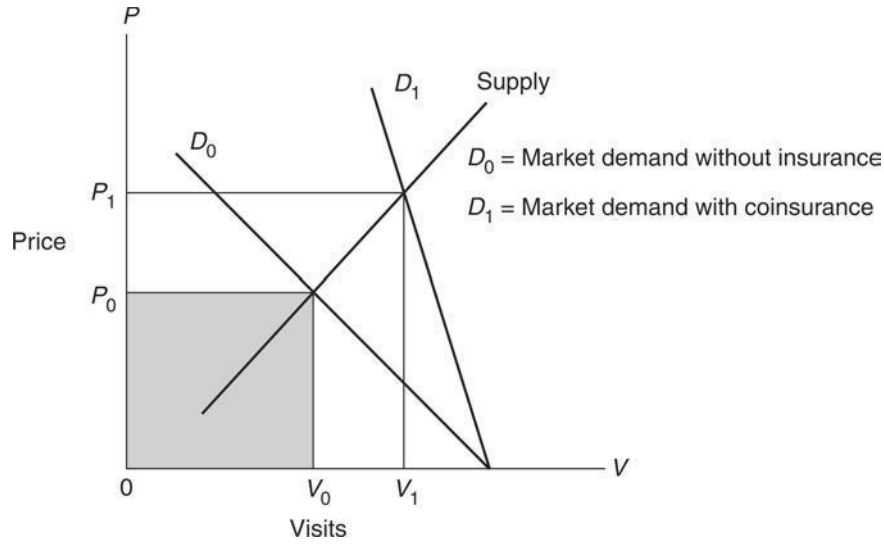


Figure 9.8 Market Impact of Coinsurance

supply curve slopes upward, indicating that higher prices might be required to motivate producers to offer greater market quantities.

Figure 9.8 shows an equilibrium of price and quantity with an upward-sloping supply curve. The original market equilibrium price is P_0 and the equilibrium utilization is V_0. In this case, if coinsurance rates are generally reduced, the increased market demand will raise market quantity demanded to V_1 and the market price to P_1. Total health care expenditures will rise from P_0V_0 to P_1V_1. Many economists feel that such effects are major reasons for the increases in health care costs in the United States.

Issues in Measuring Health Care Demand

With the current concern about health care expenditures, reliable estimates of demand elasticity become essential. Recall that elasticity measures the responsiveness of demand to a change in a related variable. Price elasticity helps determine the effects of health insurance practices and policies. The effects of public policies to improve the accessibility of health care will depend on the money and the time price elasticities.

Prices, incomes, time prices, and coinsurance rates do matter. Increased prices and coinsurance rates reduce demand for services. Raising income increases the demand for services. "Economic" factors are not consumers' only considerations, but they matter.

In this section, we focus on variables of interest to science and public policy. We examine how health care demand responds to money price, insurance coverage, and time price. In addition, we examine the effects on market demand of income and other variables. Each study attempts to apply econometrics to estimate a demand function statistically; all variables relevant to demand are considered simultaneously. For ease in exposition, we consider the important variables separately.

Reviewing the difficulties faced by researchers and the differences between studies can be helpful to understanding the results. We ask why the reported elasticities vary so often from one study to another? In most cases, the differences arise because of the different choices the researcher made in the face of problems common to research in this field. We identify five issues.

Individual and Market Demand Functions

Our analysis so far has focused on the individual. It suggested the following type of demand function for physician visits, referred to as V:

$$V = f\,(P, r, t, P_0, Y, \text{HS}, \text{AGE}, \text{ED}, \ldots)$$

where P is price per visit, r is the patient's coinsurance rate, t is a time price, P_0 is the price of other goods, Y is a measure of income, HS is the patient's health status, and AGE and ED stand for variables such as age and education to reflect other need and taste factors. This functional notation shows that certain variables are likely to affect V, but it does not specify the relation exactly.[1]

Often, however, economists wish to look at market demand functions. Care is needed to move from individual to market demand. Even the measure of utilization poses a challenge. For example, most studies will use the number of visits per person (rather than the total quantity of visits) as the dependent variable. They then attempt to control for the size of the market by considering total population. This leads to serious problems in the interpretation of results.

Measurement and Definitions

Unlike the carpenter's simpler problem of measuring the length of a wall, there are alternative definitions of health care quantities, as well as many alternative measuring tools. Investigators often measure the quantity of services in dollar expenditures. One problem is that expenditures reflect a complex combination of price of care, quantities of care, and qualities of care.

Alternative measures include quantity of visits, patient days, or cases treated, yet these do not necessarily measure the intensity of care. One person may spend five days in the hospital for observation; another may spend five days for brain surgery. Consequently, the literature contains a variety of measures and a variety of reported elasticities.

It is also difficult to define the price of services. Because of the prevalence of health insurance, most patients do not pay the full price for their treatments. Moreover, the price they pay may be related to the size of the bill because of deductibles, coinsurance, or limits. A $50 treatment, for example, may cost $50 if it occurs before the deductible limit is reached, or $10 (assuming 20 percent coinsurance) if it occurs after the deductible limit has been reached. The statistical problems in this case are fairly complicated, but it suffices here to note that the resulting price elasticities may vary.

Differences in the Study Populations

Different researchers, naturally, use different samples or populations. Elasticities will differ between populations and even within populations at different points in time. For example,

many believe that income elasticities for health care have become smaller over the years in the United States, presumably because of the effects of programs like Medicare and Medicaid.

Furthermore, it is possible, in theory, for Californians to have a different price elasticity for physician services than Minnesotans. People in one state may be older (for example, Florida) or have better access to larger varieties of health providers. It is theoretically possible that people will exhibit different price elasticities for dental care than for pediatric care. Thus, some variation in reported elasticities is inevitable even when one uses the "same" measures, definitions, and techniques.

Data Sources

Populations differ between studies, and the data sources may differ in ways that result in different elasticity estimates. For example, a common source of health care data is the insurance claim. Claims data, however, are limited to services covered by insurance and used by the insured. Furthermore, claims data often lack detail on individuals' characteristics, such as education and income. In contrast, health interview survey data often incorporate personal data, but their accuracy depends on the recall ability of the people being interviewed. Databases such as the Medical Expenditure Panel Survey (MEPS), collected by the Department of Health and Human Services, now provides valuable up-to-date data archives for policy analyses (http://meps.ahrq.gov/mepsweb/).

Experimental and Nonexperimental Data

Much of health care demand research used nonexperimental data, and thus the researcher could not control the environment or assure that other extraneous variables were held constant. These data typically represent samples across individuals or markets—that is, a slice of experience. If the necessary assumptions hold, then available statistical techniques can provide valuable analytical insights.

A natural experiment, for example, may occur when a given area changes its health insurance plan (e.g., Tilford and colleagues (1999) studied the response of previously uninsured school children in the Mississippi Delta region of Arkansas after a program provided them with health insurance). The change enables one to observe differences in health care utilization before and after. We presume that only the policy changes; all other factors are held constant. Unfortunately, other demand-related factors often change also.

In a controlled experiment, subjects are randomly assigned to treatment and control groups to measure responses directly to changes in the levels of demand-related variables (Box 9.2 describes a unique opportunity that created a natural experiment). Such experiments, however, are generally costly to perform and are not without their own difficulties.

BOX 9.2

Oregon's Health Insurance Experiment

In 2008, the state of Oregon expanded its Medicaid program, to aid low-income households, in a way that enabled analysts to perform randomized controlled studies. Oregon received permission from the federal government to create a lottery that allowed approximately 30,000 low-income adults, from a waiting list of about 90,000, to enroll in Medicaid if they met the eligibility requirements. Together with survey data

from those on the list, the lottery allowed for comparisons between those who were selected (treatment group) and those who were not selected (control group) across a wide range of economic and health outcomes indicators.

In the first year, the treatment group had substantially higher rates of utilization, including primary and preventive care as well as hospitalization; lower out-of-pocket costs; and better self-reported physical and mental health (Finkelstein et al., 2012). With the exception of physical health, many of these results persisted after two years (Baicker et al., 2013). However, Baicker and colleagues found that Medicaid coverage had no significant effects on the prevalence or diagnosis of two important health outcomes—hypertension and high cholesterol levels. Among the most surprising results, those with Medicaid coverage substantially increased emergency department visits by about 40 percent relative to the control group (Taubman et al., 2013). Emergency use increased even among the kinds of visits, e.g., "primary care treatable," that could be handled in outpatient settings.

Empirical Measurements of Demand Elasticities

Price Elasticities

Health care demand studies focus on price elasticity. Table 9.2 reports a selection of estimates by type of care. The dependent variable in each case is the quantity demanded. In some cases, it is a market aggregate, such as admissions per capita; in other cases, the unit of observation is the individual consumer. Most reported elasticities range between 0.0 and –1.0, indicating that consumers, while responsive to price, are not responsive to a substantial degree. Suppose that the price elasticity for physician services was between –0.08 and –0.18, and physicians raised their prices by 10 percent. This would reduce consumption by 0.8 to 1.8 percent.

Table 9.2 Price Elasticities from Selected Studies

Study	*Dependent Variable*	*Price Elasticity*
All Expenditures:		
Manning et al. (1987)	All expenditures	–0.17 to –0.22
Physician Services:		
Newhouse and Phelps (1976)	Physician office visits	–0.08
Cromwell and Mitchell (1986)	Surgical services	–0.14 to –0.18
Wedig (1988)		
Health perceived excellent/good	Physician visits	–0.35
Health perceived fair/poor	Physician visits	–0.16
Chandra et al. (2010)	Physician visits	–0.10

continued

Table 9.2 *continued*

Study	*Dependent Variable*	*Price Elasticity*
Hospital Services:		
Newhouse and Phelps (1976)	Hospital length of stay	–0.06
Manning et al. (1987)	Hospital admissions	–0.14 to –0.17
Nursing Homes:		
Chiswick (1976)	Nursing home residents per elderly population	–0.69 to –2.40
Lamberton et al. (1986)	Nursing home patient days per capita elderly	–0.69 to –0.76
Pharmaceuticals:		
Chandra et al. (2010)	Prescription drugs	–0.08 to –0.15

Table 9.3 Firm-Specific Price Elasticities

Study	*Dependent Variable*	*Price Elasticity*
Physician Services:		
Lee and Hadley (1981)	Physician price	–2.8 to –5.1
McCarthy (1985)	Physician visits	–3.1 to –3.3
Hospital Services:		
Feldman and Dowd (1986)	Hospital patient days	–0.7 to –0.8
	Hospital admissions	–1.1
Gaynor and Vogt (2003)	Hospital discharges	–4.9
Nursing Homes:		
Mukamel and Spector (2002)	Case-mix adjusted days	–3.5 to –3.9

A further distinction among studies should be made. The price elasticities reported in Table 9.2 measure the consumer's or the market's response to price changes. That is, they do not relate to a particular seller, but instead represent the demand for the health care good or service in general.

The demand for physician care in the market will be less elastic than the demand for the services of a particular physician. For example, suppose a medical symptom is worrisome, and the patient chooses to see a physician. The more worrisome the symptom, the less responsive he or she is likely to be to market price. Which physician to see is a completely different question. The consumer who knows the quality and price of each physician would choose the lowest-priced physician among those of equal quality. The point is that there are few substitutes for physician care, but there are many substitutes among individual physicians. Thus, firm (physician)-specific demand will be more price responsive than overall demand.

This is illustrated in Table 9.3, which reports studies of firm-specific demand elasticities. As we observe, these elasticities tend to be considerably higher in absolute value than most of the elasticities reported previously in Table 9.2.

The firm-specific elasticities have further significance. They indicate the degree of competition in the health services market. Under perfect competition, firm-specific elasticities will approach negative infinity because consumers will respond to a firm's price increase by instantly going to a competitor. The reported physician care elasticities may be large enough that competition is a reasonable approximation. In contrast, the smaller elasticity estimates for hospitals (that is, closer to 0) suggest considerable market power.

Individual Income Elasticities

Economic theory suggests that increased income causes increased purchases for most goods. Most goods have positive income elasticities and are referred to as normal goods. Those with negative elasticities are referred to as inferior goods. Table 9.4 reports estimated income elasticities for a selection of studies by type of health care. In most cases, the magnitudes are small. This indicates that while health care is generally a normal good, the response is relatively small; that is, inelastic.

Income elasticities also help define when goods are necessities or luxuries. We call goods "necessities" when the income elasticity is between 0 and +1. When income elasticities exceed +1, goods are "luxuries." From the properties of elasticities, a 1 percent rise in income increases the budget share devoted to a luxury and decreases the budget share devoted to a necessity.

From Table 9.4, the results are not surprising; people commonly perceive health care to be a necessity.

Income Elasticities across Countries

Given these findings, it may be surprising that studies of aggregate health expenditures across countries report substantially higher income elasticities. Often the magnitudes exceed unity. An early cross-national study was published by Newhouse (1977) who regressed per capita medical expenditures for 13 developed countries on a variable representing per capita income. From the estimated coefficient of this equation, he then calculated the implied income elasticity for various levels of income. The elasticity estimates

Table 9.4 Income Elasticities from Selected Studies

Study	*Dependent Variable*	*Income Elasticity*
All Expenditures:		
Rosett and Huang (1973)	Expenditures	0.25 to 0.45
Hospital Services:		
Newhouse and Phelps (1976)	Admissions	0.02 to 0.04
Physician Services:		
Newhouse and Phelps (1976)	Visits	0.01 to 0.04
Nursing Homes:		
Chiswick (1976)	Residents per elderly population	0.60 to 0.90

ranged from 1.15 to 1.31. He concluded that despite within-country results showing health care to be a necessity, health care in fact is a luxury good.

Parkin and colleagues (1987) pointed out several potential weaknesses in most existing cross-national studies, but despite their objections, offered improved results that tended to support the finding of cross-national income elasticities greater than 1.0. Gerdtham et al. (1992) and Getzen and Poullier (1992) also lend support to the result.

Is it inconsistent that within-country health care income elasticities are small, while cross-national estimates exceed 1.0? Can health care be a necessity at the individual and market levels but a luxury at the country level? A hypothetical example illustrates that income elasticity results at the national aggregate level do not necessarily apply to individual or market level. Suppose that two countries, one rich and one poor, each provided free health care to their citizens irrespective of income. Then within-country income elasticities might be small if not zero. Yet the richer country might provide greater quantities, higher technology, and better qualities of health care to each of its citizens. Thus, the cross-country income elasticities could be high.

These and related ideas are more fully developed by Getzen (2000), who shows that the individual's response to more income is different than the nation's response to more income. He also notes that symptoms of illness and pain are often more important reasons we as individuals seek out the doctor, while the available health care resources and technologies at the national level often reflect the nation's economic well-being.

The results of this small but well-established line of research have been challenged by two Canadian researchers, Blomquist and Carter (1997). By studying a large set of countries over time, observing time patterns and country-specific effects, they tentatively concluded that health spending grows about 2 percent faster than income in a manner suggesting the role of technological progress. But what about the original research goal; are the income elasticities for health spending greater than 1.0? Of 18 countries studied, they find that 11 income elasticities were either less than 1.0 or so close to 1.0 that the null hypothesis (that the elasticities equaled 1.0) could not be ruled out. Whatever direction this line of research work takes in the future, researchers are gaining increasingly sophisticated understanding of the methods and challenges of performing cross-national studies.

Insurance Elasticities

Consumer responses to changes in insurance are important because insurance coverage has grown dramatically in the past 50 years and because we frequently must consider possible changes in social insurance. While the issue of insurance effects must be treated separately from price effects, they are closely intertwined with the issue of price elasticities.

Consider a health insurance policy where the consumer pays a fixed percentage of the bill—that is, a fixed coinsurance rate, r. In such a case, the net price that the consumer pays would be a simple multiple of the market price, P:

$$\text{Net price} = rP$$

When the market price increases by 1 percent, so does the net price; that is:

$$1.01 \times \text{Net price} = r\,(1.01P)$$

Under such an insurance plan, the coinsurance elasticity would be the same as the price elasticity.

However, most health insurance plans are not so simple. In practice, they include deductibles and maximum dollar expenditure (MDE) limits in addition to the coinsurance rate. The result is that the effective coinsurance rate depends in part on the size of the bill. In practice, price and coinsurance elasticities will differ somewhat.

A further difficulty arises because most studies have examined nonexperimental data. One of the major concerns with nonexperimental data is that the groups compared are not always randomly selected. For example, suppose that a company allows its employees to enroll in either a high-coverage plan or a low-coverage plan. Some people may choose to work for the company because it offers the high-coverage insurance plan. Others who expect to use large (low) amounts of services naturally enroll in the high- (low-) coverage plan. Still others, if sufficient insurance is not available from the company, will purchase more generous insurance privately. If the demand analysis proceeds by comparing these groups, the results may misstate the true effect of coinsurance. This is because the major decision was made in deciding which group to join. The behavior is known in economics as *adverse selection*.

The RAND Corporation, funded by the United States Public Health Service, mounted an experiment beginning in 1974. Known as the RAND Health Insurance Experiment (RHIE) and led by Joseph Newhouse, the study randomly assigned households at six sites across the nation to groups that had different levels of cost sharing, ranging from free care to care with 95 percent coinsurance and including a maximum dollar expenditure limit. The families were paid a lump-sum payment to be sure that no family was made worse off by the experiment. Because the assignment was random, adverse selection could be minimized, and the random assignment of coinsurance also allowed researchers to investigate the effects of coinsurance on expenditures.

They observed family health care use and expense experience over a period that varied from three to five years for various experimental groups. This intensive and expensive experiment improved our understanding of the response of health care consumers to economic incentives.

Coinsurance has a considerable effect on the level of average medical expenditures. From an extreme of a 95 percent coinsurance to the opposite extreme of free care, or zero coinsurance, the average family's medical expenses increase by nearly 50 percent, from $679 to $982. Even hospitalization rates are responsive, increasing from 7.9 percent of those in the 95 percent coinsurance group to 10.3 percent in the free care group, representing an increase of about 30 percent. The RAND experimental data and analysis show that both price and insurance do matter considerably. Newhouse and colleagues (1993) conclude:

> All types of service—physician visits, hospital admission, prescriptions, dental visits, and mental health service use—fell with cost sharing. There were no striking differences among these services in how their use responded to plan. Another partial exception was demand for mental health services—which, the results indicate, would have been more responsive than other services to cost sharing had there been no cap on out-of-pocket expenditure.
>
> (pp. 338–339)

The RHIE was not designed to track the effects of insurance on the elderly. Other researchers, however, find that the elderly consume more health care when they are more fully insured. The elderly, who are automatically eligible for Medicare hospital insurance, may opt for additional coverage. A confounding factor, one which the randomized

experimental approach of the RAND study eliminates, is the possibility that those who buy the extra insurance might be those who expect to be more ill, another form of adverse selection. Hurd and McGarry (1997) separated out this confounding issue, and they conclude that the insurance effect among the elderly is due primarily to the way in which insurance changes the economic incentives that accompany illness rather than adverse selection into the insurance pool.

Finally, although the RAND study is often considered the methodological gold standard, it was conducted nearly 40 years ago. Much has changed in the health economy since then, especially the growth of managed care. Meyerhoeffer and Zuvekas (2010) use comprehensive annual surveys of the U.S. civilian population over 1996–2003 to estimate more recent price elasticities for physical and mental health care. Elasticities for both services were low but, surprisingly, the price elasticity of demand for mental health visits (–0.05) was even lower than ambulatory visits for physical health problems (–0.12).

Impacts of Insurance on Aggregate Expenditures

RAND researchers, estimating coinsurance and income elasticities to be approximately 0.2, sought to calculate the demand-related portion of the post-World War II real increase in U.S. health expenditures due to the spread of health insurance. The answer was "not much"—only 10 percent of the increase.

Using the RAND income elasticity of 0.2, the post-war income increase accounted for about another 10 percent. Therefore, according to the RAND investigators, coinsurance and income accounted for about one-fifth of the total increase in real health expenditures. Subsequent research (Peden and Freeland, 1998) determined that about half of the expenditure increase was due to induced technological innovation. Those authors also attributed a higher impact (20 percent) to increased income.

Other Variables Affecting Demand

The studies we have reviewed often incorporate many other variables of interest in the demand function estimates, and considerable information relevant to policy issues has been obtained.

Ethnicity and Gender

Many studies of demand examine the influence of race, and find that blacks tend to consume less medical care than the other large, self-identified ethnic groups when other factors are held constant. Because the disparities in utilization across racial and ethnic groups have been so large and persistent over time, in 1999 Congress mandated the Agency for Healthcare Research and Quality to publish an annual disparities report. (See Box 9.3 for further discussion.) Although the majority of studies of ethnic differences in health care have focused on the experience of blacks and Hispanics, other ethnic differences also have been noted, often when a given disease, for example Tay-Sachs disease, appears predominantly within one group—in this case, Jews of Eastern European origin.

BOX 9.3

Disparities in Health Care: A National Priority

Disparities across racial, ethnic, and socioeconomic groups in health outcomes and health care utilization are well-documented. The *Healthy People 2010* (U.S. Department of Health and Human Services, 2000) initiative placed the elimination of disparities on the national agenda. Yet, the most recent report from the Agency for Healthcare Research and Quality (2015) covering 2012 indicates that disparities remain common. The Agency uses over 250 measures of quality (e.g., pregnant women receiving prenatal care in the first trimester) and of access (e.g., people who have a specific source of ongoing care). Those in poor households had worse access than those in high-income households on all the access measures; the poor also received lower quality of care on more than half the quality measures. Blacks had worse access than whites on half the measures and they received worse care than whites on about one-third the quality measures. More disturbing, the Agency found that most access and quality disparities related to race, ethnicity, or income have not changed significantly over time despite the national attention and policy priority given to this problem.

Why? Is there discrimination in health care delivery against certain population groups as some have suggested? There are no easy answers but the Institute of Medicine's report to Congress on the extent and sources of the disparities (Smedley, Stith, and Nelson, 2002) greatly raised awareness of the complexity of the underlying issues. The report recognized that differences in access to care are major contributors to disparities in utilization and health outcomes, but also that there are many other confounding factors including discrimination and differences in preferences and propensities to seek care across groups. Economists have sought to develop methods that distinguish among the various sources that account for disparities (Balsa, Cau, and McGuire, 2007; David and Harrington, 2010; and Mahmoudi and Jensen, 2012).

Part of the differences may be of social origin; for example, blacks and other ethnic groups may be reluctant to seek care or may be treated differently by white physicians. However, ethnic differences, more broadly understood, may help explain geographic patterns across the United States. For example, Westerners tend to rely less on hospitals and physicians, while residents of the more snowbound north central states rely more heavily on nursing home care for the elderly than other areas on a per capita basis.

Females differ from males most clearly in their time pattern of medical care usage. During childbearing years, women are relatively heavy users of health care, but women are healthier in the long run and they predominate in the numbers of the elderly, and thus among physicians' older patients. Though a great deal of public attention and concern in the past decade has turned to the science of treating diseases prominent among women, death rates for cancers are often as high or higher among men as among women. For example, mortality rates from prostate do not differ much from mortality rates from breast cancer. Myocardial infarction (heart attack) is a notorious killer of men, though women's rates have been increasing.

Thus, researchers will continue to study the differences in medical demand between the sexes, and the many differences among ethnic and cultural groups. These include not just differences among skin color groups but among the many subcultures and local cultures within these larger groups. These differences may offer the explanation for demand variations not accounted for by the usual demand variables and may yet help identify and explain many health demand questions that have remained unanswered.

Variations in sexual behavior have proven tremendously important in explaining variations in the pattern of infections with HIV and mortality due to AIDS. Homosexuals in the United States were long the major group at risk and most prominent among deaths, with intravenous drug users second in numbers. Though much public effort was addressed to the prevention of a feared epidemic among heterosexuals, especially youth, the pattern has remained steady. In contrast, the world's attention is turned more toward Africa, where transmission of the disease is primarily heterosexual, and the size of the epidemic has formed a crisis for world public health efforts.

Urban versus Rural

Studies sometimes find differences in health care usage due to rural status. If rural residents use less care, the reasons why are not necessarily clear. Rural dwellers may differ culturally, and some analysts argue that this factor is more important to one's perception of life than ethnicity is. Whether born to rural life or to have adopted it, it may become linked to tastes, health status, and relative reluctance to seek out a physician. The lesser health demand by Westerners, already identified previously, could be understood in this view as an artifact of the predominance of rural areas in the region.

The contrasting argument is made that the greater travel distances required to obtain health care in rural areas, rather than rural culture or tastes, account for the demand patterns. Thus, studies of geographical patterns of health care demand must take special care to measure the full price of physician or hospital care, that is, to include the travel time price.

Education

Education is strongly associated with better health. If you are a college student, the odds are very good that you are healthier than your noncollege counterparts. As in the demand for health capital model, this may be because you are a more efficient producer of health, you are less likely to smoke, and you are more likely to eat a healthful diet. Or it may be that you are the sort of person with a long-term goal, and to meet that goal you have identified the need to take proper care of yourself. For the researcher sorting out such questions, the complexity of the issue is multiplied by confounding factors, especially income. Educated people tend to earn more, a fact not lost on most college students. We then must determine whether education improves one's health, or whether the income it brings affords a healthier life.

Age, Health Status, and Uncertainty

Older people consume three to four times more health care than the younger population. Though the relationship is no surprise, Grossman's theory of this pattern makes the issue more intriguing to the research community. If we invest in our health each period and yet our health depreciates somewhat during each period, why do we necessarily buy greater quantities of health care as we get older? Very plausibly, as Grossman assumed, the depreciation rate increases as we age, thus greater inputs of our own time and health care are required

to restore our health. This makes the correlation of health demand and age appear logical. Perhaps more interesting is the relationship of health status to the price elasticity of health care demand. Theory is not clear on this point, but it is plausible that sicker people will tend to be less sensitive to price.

Wedig (1988) finds that the price elasticity of the decision to seek health care tends to be lower in absolute value for those with poorer health status, regardless of which measure is used to record health status. However, no clear pattern over health status can be determined with respect to level of care—that is, the amount of health care consumed given that the consumer has chosen to seek health care.

Finally, uncertainty will affect health care demand. When a consumer, worried about a future health risk, seeks advice or preventive treatment, we call this a precautionary demand (Picone, Uribe, and Wilson, 1998). Elderly patients, for example, may smooth their utility over time by spending now to avoid sharp drops in well-being and mobility in the future. Some empirical evidence suggests that older people have somewhat less tolerance for risk and that one's degree of risk tolerance influences one's decision whether to buy health insurance.

Conclusions

Demand theory is crucial to our understanding of health care markets. The substantial increases in out-of-pocket costs for prescription products experienced by many patients have affected utilization of drugs in the expected negative direction (see Chapter 17 for specifics). Hospitals and other providers continue to compete for patients as well as for contracts with managed care organizations. The more recent estimates of price elasticities for hospitals and nursing homes shown in Table 9.3 indicate that the demand facing both types of providers is even more sensitive to price than prior studies have shown. Time and distance can also be important as theory suggests. In a dramatic demonstration, Currie and Reagan (2005) found that each additional mile to the nearest hospital reduces the probability that central-city black children have a checkup by 3 percentage points, regardless of whether the children are privately or publicly insured.

An analysis of the demand for physician care in 12 European Union countries illustrates the universal relevance of demand theory. Jiménez-Martin and colleagues (2004) show that one-third to one-half the variability in demand across countries is explained by differences in age, income, and the physician's role in the health care system. Other factors include whether the general practitioner (GP) acts as a gatekeeper and whether physicians are capitated, salaried, or paid on a fee-for-service basis. In fact, the frequency of GP visits increases and the probability of contacting specialists as well as the number of visits to specialists decrease in countries where GPs are gatekeepers. Such results can help policymakers design reforms that better meet their efficiency and cost targets.

Reliable estimates of elasticities for specific services can also inform policy in other ways. For example, some reforms propose higher copayments for more discretionary and potentially lower-value care. Using expenditures in private insurance markets in Chile, a middle-income country, Duarte (2012) estimated elasticities ranging from near zero for urgent care (appendectomy) to –2.08 for highly elective care (psychologist visits).

Finally, a good understanding of demand theory serves as the rationale for market-based, consumer-driven approaches to health system reform. Under health reform legislation in the United States, the future of market-based strategies remains unclear. Nevertheless, nearly 25 percent of all covered workers in 2015 were enrolled in consumer-directed and other high-deductible health plans. Described more fully in Chapter 22, these plans typically

involve high-deductible catastrophic insurance and other features that enable consumers to take greater control over their spending decisions. By "empowering" consumers, supporters of this strategy envision a more competitive system in which the decisions of cost-conscious patients restrain fees and limit use of marginally beneficial care. It is still too early to be able to evaluate fully the impact of high-deductible plans, but their growth has created opportunities for applications of demand theory. Box 9.4 describes one of the major challenges to patients with such plans.

BOX 9.4

How Much Will That Hospitalization Cost Me?

The success of high-deductible plans and other market-based initiatives relies on patients' abilities to make rational choices based on price and quality. Health care prices are hardly transparent to patients. Hospital prices are especially troublesome, in part, because patients may be billed separately by the hospital, and various physician providers such as the surgeon, the radiologist, and the anesthesiologist.

Rosenthal, Lu, and Cram (2013) investigated the challenges patients face when they try to obtain pricing information. Using total hip arthoplast (THA), a common orthopedic procedure among the middle-aged and elderly, the authors telephoned two hospitals (up to five times if necessary) in each state and the District of Columbia, as well as 20 leading orthopedic hospitals, to determine the "bundled price" that includes all services for a 62-year-old uninsured grandmother who was willing to pay out-of-pocket. Nine of the 20 top-ranked hospitals provided the bundled price but only 10 of the 102 other hospitals were able to do so. After separately contacting hospitals and physicians, the authors obtained complete bundled prices from three additional top ranked hospitals and 54 other hospitals.

Aside from showing just how difficult it can be for patients to obtain meaningful price information, the study revealed extraordinary variation in the bundled prices. Prices at top-ranked hospitals varied by more than 8-fold, ranging from $12,500 to $105,000 with a mean of $53,140. Prices at the other hospitals varied by almost 12-fold, ranging from $11,100 to $125,798 with a mean of $41,666. Although there can be substantial savings for those who "shop around," the lack of price transparency, combined with the need to find a physician with admitting privileges to lower cost hospitals, and the costs of travelling to the hospital, can make it difficult to actually attain the savings.

Summary

1 The theory of rational choice over health care and other goods helps explain our decisions because many health care options are not urgent, leaving room for thoughtful consideration or at least some planning.
2 In addition to consumers' decisions, physicians serve as the patient-consumer's agents and can make rational choices even in urgent situations.

3 Depicting the consumer's choice requires knowing preferences, as described by a set of indifference curves, and resource constraints, described by the budget line indicating income and market prices.
4 Consumer equilibrium occurs only if the rate at which they are willing to trade two goods, or MRS, equals the price ratio at which they are able to trade the two goods. In equilibrium, a dollar buys the same marginal utility from all goods.
5 Price elasticity, E_p, is the ratio of the percent change in quantity demanded to the percent change in price. Income elasticity, E_Y, is the percent change in quantity demanded divided by the percent change in income.
6 The time spent acquiring services constitutes a substantial portion of the economic costs. The discrepancy between the total economic prices (including time) and the money prices will be especially large for low-priced services, services with small patient copayments, and for patients with high time costs.
7 Insurance plays a major role in health services demand. Many health care purchases are at least partially covered by health insurance so that a portion is paid for by someone other than the consumer.
8 The impact of coinsurance depends critically on the price elasticity of demand for health care. If consumers do not respond to price changes in the absence of insurance, changes in coinsurance will have no impact on quantity of services demanded.
9 Coinsurance makes the demand curve for health services less responsive (less elastic) with respect to the price.
10 Quantity of services is often measured by dollar expenditures. One problem is that expenditures reflect a combination of price of care, quantity of care, and quality of care. Alternatively, quantity may be measured in numbers of visits, patient days, or cases treated.
11 It is often difficult to define prices of services since insured patients usually do not pay the full price. Moreover, the net price paid by consumers is influenced by deductibles, coinsurance, or other limits.
12 Most reported price elasticities indicate that consumers respond to price changes. However, these elasticities (between 0.0 and –1.0) are not large compared to many other goods and services.
13 In most cases, income elasticities are low. While health care is a normal good, since its demand increases with income, the response is relatively small. However, at aggregate levels, across countries income elasticities often exceed +1.0.
14 Coinsurance has a considerable effect on the level of average medical expenditures. Both price and insurance matter.
15 Income and insurance changes since World War II may explain approximately one-fifth of the increase in U.S. health expenditures, through increased demand for services. Researchers attribute much of the remainder to increased costs brought on by technological change.

Discussion Questions

1 Discuss how time costs affect health care demand, and speculate on this and possible other reasons for the lower observed per capita demand for health care in the western United States.
2 Define *price elasticity of demand*. How does an increase in the coinsurance rate affect the consumer's price elasticity?

3 Why are firm-specific demand price elasticities higher than elasticities for demand in general? Why does a high elasticity indicate a very competitive market?
4 For the following pairs of services, which of the two services would you expect to be more income elastic? More price elastic?
   (a) Surgical services versus allergist services.
   (b) Heart surgery versus cosmetic surgery.
5 It has been discovered that countries with higher per capita incomes spend more than proportionally as much on health care. What does this imply about the cross-national income elasticities? Why might this occur, even though individual income elasticities seem to be quite low?
6 The frequencies of health care visits are often used to measure service demand. Many, however, criticize the use of this variable. What are some pros and cons of the use of visits?
7 We often speak of how price rations goods. What are other rationing measures in clinics in which free care is provided?
8 Explain or show why the impact of changes in coinsurance rates on demand depends on the elasticity of demand. What sorts of health care goods or services will be responsive to changes in coinsurance rates? What sorts will tend to be relatively less responsive?
9 A profit-maximizing firm, finding that its demand is inelastic, will necessarily find it profitable to increase its price; therefore, its equilibrium price elasticity will necessarily be greater than 1.0 in absolute value. Are the market- and firm-specific elasticity data reported here consistent with this theory?
10 The consumer's indifference curves in Figure 9.2 indicate substitutability between visits and other goods. What will the indifference curves look like if the consumer perceives no substitutability? What will happen to the elasticity of demand in this case?
11 Some argue that wide disparities in utilization rates across racial and ethnic groups are indicative of discrimination (see Box 9.3). Use indifference curve analysis to explain why it may be difficult to distinguish between discrimination and differences in socioeconomic factors such as incomes and preferences.
12 Box 9.4 describes the difficulty that consumers may have in searching for the total "bundled" price for their hospital care. What are some of the implications of this lack of pricing transparency for (i) the individual consumer, and (ii) the efficiency of health care delivery?

Exercises

1 Suppose that Martha's income is $40,000 per year. She can spend it on health care visits, which cost $80 per visit, or on groceries (standing for all other goods), which cost $100 per bag of groceries. Draw Martha's budget constraint. Using indifference curves, show Martha's optimum if she buys 300 bags of groceries per year.
2 Suppose that Martha's income rises to $42,000 per year, and that she increases her consumption of health care visits by five visits. Using the graphs for Exercise 1, draw the new equilibrium. What is her income elasticity of demand for health care visits?
3 Consider the following information on Alfred's demand for visits per year to his health clinic, if his health insurance does not cover (100 percent coinsurance) clinic visits.
   (a) Alfred has been paying $30 per visit. How many visits does he make per year? Draw his demand curve.

(b) What happens to his demand curve if the insurance company institutes a 40 percent coinsurance feature (Alfred pays 40 percent of the price of each visit)? What is his new equilibrium quantity?

P	*Q*
5	9
10	9
15	9
20	8
25	7
30	6
35	5
40	4

4 Suppose that a consumer makes V_0 physician visits each year at a price of P_0. If the price elasticity is –0.4, what will happen to the number of visits if the price increases by 10 percent? What will happen to total physician expenditures? Why?
5 If the price elasticity of demand is –0.5 and the income elasticity is +0.3, then what will be the effect of a simultaneous 10 percent increase in price and a 10 percent increase in income on health expenditures?
6 Draw a diagram for hospital care that reflects the income-elasticity estimates found empirically. As income increases, what happens to the proportion of income spent on hospital care?
7 Would the opportunity cost of waiting time be higher for higher-income people or lower-income people? Given your answer, for which income group would money price tend to be a smaller portion of the full price?
8 Explain how the demand for health insurance is related to the demand for health care. Would the demand for health care then depend also on whether the person paid for the insurance or alternatively was provided the insurance at a subsidized cost?

Note

1 Econometricians often use the OLS method discussed in Chapter 3. In this case, the regression is: $V = b_0 + b_1P + b_2r + b_3t + b_4P_0 + b_5Y + b_6\text{HS} + b_7\text{AGE} + b_8\text{ED} + \varepsilon$ with the variables defined as before, and ε is the error term.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 10

Asymmetric Information and Agency

In this chapter

- Overview of Information Issues
- Asymmetric Information
- Application of the Lemons Principle: Health Insurance
- The Agency Relationship
- Consumer Information, Prices, and Quality
- Conclusions

The traditional theory of demand, as we have seen, begins with the assumption that individuals are fully informed about prices, quantities, and the relationships of medical care and other inputs to their levels of health. We examined decision making within a model that assumed perfect information. Depending upon the model's purpose, such an assumption may be justified even if it is not realistic. However, a more complete understanding of the health economy requires particular insight into the effects of various informational problems in health care markets.[1]

Though most have long regarded imperfect information as a feature of the health economy, its specific effects were not well understood until the economics of information emerged as a distinct specialty. The development of tools to study asymmetric information and agency relationships has greatly enhanced the field of health economics. Asymmetric information encompasses situations where buyers and sellers have different levels of information; agency concerns situations where, for lack of information, buyers or sellers rely on other parties to help make decisions.

Overview of Information Issues

The markets for many health care services and for insurance in particular exhibit significant degrees of asymmetric information and agency relationships. For example, adverse selection, a phenomenon in which insurance attracts patients who are likely to use services at a higher than average rate, results from asymmetric information. Most agree that potential beneficiaries have better information than the insurer about their health status and their expected demand for health care. As a result, premiums for higher-risk patients will be underpriced, encouraging such patients to overinsure, whereas the opposite holds true for lower-risk patients. Adverse selection reduces the efficiency of health insurance markets while redistributing income from the healthy to poorer risks.

Information and agency problems account for many other important characteristics of health care markets. The possible preference for health care delivery by nonprofit hospitals and nursing homes (Chapter 13) has been attributed to patients' lack of information and inability to discern quality. For some patients, a nonprofit status might reassure clients of the higher quality because decisions are independent of a profit motive. Lack of quality information also is an important motive for licensure and other regulatory measures.

The present chapter has three goals. First, we introduce information asymmetry, describe its relative prevalence, and determine its consequences, especially for insurance markets. It will quickly become clear that adverse selection in insurance is only one consequence of asymmetric information.

Asymmetric information, as when a patient is less well informed about appropriate treatments than the attending physician, typically leads to an agency relationship between the patient and provider. The second goal of this chapter is to describe the agency relationship and examine some of the problems arising in health care markets from imperfect agency. We look at the special and controversial case of supplier-induced demand (SID) as an asymmetric information/agency problem in Chapter 15.

Finally, we seek to examine the effects of imperfect consumer information on the price and quality of health care services. Despite consumers' informational disadvantages, they often influence markets in predictable ways. Here and throughout the chapter, we identify arrangements that commonly evolve to reduce the disadvantages for the less well-informed parties. Thus, the ultimate consequences of asymmetric information and imperfect agency on the efficient functioning of markets are often less severe than one might initially assume.

Asymmetric Information

Basic microeconomic theory usually includes an assumption that the market being analyzed exhibits perfect information. Under conditions of perfect information, all consumers and producers have complete information on all prices, as well as the quality of any good or service available in the market. Consumers also will be as well informed about the product as the seller.

Although information is never perfect in the real world, perfect information serves as a useful starting point because the properties and predictions of the standard models relying on this assumption are so well understood. Also, as students of economics are repeatedly taught, the relevant issue is whether the predictions derived from standard models apply to real-world markets. In many cases, the predictions derived from these models hold up reasonably well.

Cases do arise, however, where imperfect information does seem to matter. During the past four decades, economists have developed new insights into the effects of imperfect and asymmetric information. This section examines some of that work, including contributions by health economists to the specific problems of the health sector.

On the Extent of Information Problems in the Health Sector

Before investigating several contributions to the economic theory of information, we begin by asking how prevalent information problems are in the health sector. It is obvious not only that information is imperfect in health care markets, but also that information is asymmetric. Levels of information will differ among participants, such as between physicians and patients. Patients are often poorly informed compared to the provider about their conditions, the treatments available, expected outcomes, and prices charged by other providers. Furthermore, we presume that the information problems that exist in the health sector are prominent enough to require the special analysis of the economics of information.

Although we can agree that information problems arise in health care markets, we must avoid the temptation to overemphasize this point. To say that information problems exist in the health sector does not mean that these problems are necessarily worse than in any other market. Markets for insurance, other professional services, and automobile and appliance repairs also exhibit asymmetries. We should not necessarily conclude that information asymmetries in health care markets make it impossible for corrective institutions, practices, or products to evolve; nor do they necessarily preclude the possibility of competition.

Why? Pauly (1978) noted that at least half of physician visits customarily are made for services, such as general checkups or chronic care, for which the patient has some if not considerable experience. From data on the portion of medical expenditures attributable to ambulatory physician care, we can estimate that if half of this care is reasonably well informed, then about 8 percent of total medical care is informed. Reasoning in this manner about all sorts of medical care and products, Pauly concluded that plausibly "one-fourth or more of total personal health-care expenditures might be regarded as 'reasonably informed'" (p. 16). By adding nursing home services and chronic conditions, Pauly (1988a) subsequently argued that this ratio is about one-third.

Further, for several medical care issues, the provider shares in the information gap with the patient. As we will emphasize when describing the "small area variations" literature in Chapter 15, the provider often is uncertain if not uninformed about the outcomes of many medical procedures. In such cases, information asymmetry does not necessarily arise even though it may be correct that the patient is ill-informed.

Finally, economic analysts of information asymmetry problems have been able to show that markets may perform well in the face of some degree of information asymmetry provided a sufficient portion of the consumers are reasonably well informed. Perhaps a majority of consumers who use personal computers know little about their technical aspects and relative qualities and prices. However, a significant minority of consumers tends to be highly informed. In markets like this, the informed minority may be sufficient to provide the economic discipline it takes to make the market perform well so that the rest of us will find that the higher-priced computers also tend to be of higher quality.

We conclude this section by summarizing its main point. Certainly information gaps and asymmetries exist in the health sector. They are perhaps more serious for health care than for other goods that are important in household budgets. This makes it useful for the student of health economics to investigate the theory of information asymmetries and its application to health care. However, we should not overlook mechanisms to deal with information gaps. These mechanisms include licensure, certification, accreditation, threat of malpractice suits, the physician–patient relationship, ethical constraints, and the presence of informed consumers.

Will a state of relative consumer ignorance preclude high levels of competition? Will health care markets be characterized by a high degree of price dispersion and the provision of unnecessary care or care that is not in the patient's best interests? Can some of the characteristics of health care markets and the evolution of their institutional arrangements be related to asymmetric information? The following sections address these and other questions by beginning with the pioneering work on asymmetric information.

Asymmetric Information in the Used-Car Market: The Lemons Principle

Nobel Laureate George Akerlof (1970) is often credited with introducing the idea of asymmetric information through an analysis of the used-car market. Though seemingly unrelated to health care, his classic article is important for two reasons. First, it tells us much about adverse selection and the potential unraveling of health insurance markets. Adverse selection provides a key to our understanding of some major contemporary issues, such as the reasons that some may remain uninsured, or the performance of delivery systems such as health maintenance organizations. Second, Akerlof's example leads right into the issue of agency.

In Akerlof's model, used cars available for sale vary in quality from those that are still in mint condition to some that are complete lemons. Information asymmetry arises if, as is plausible, the sellers know better the true quality of their cars than do the potential buyers. Akerlof showed where such information asymmetry causes the market for used cars to perform poorly, or even to disappear entirely.

Suppose that nine used cars are to be sold (potentially) that vary in quality from 0, meaning a lemon, to a high of 2, meaning a mint-condition used car, a "cream puff." In fact, suppose that the nine cars have respectively quality levels (Q) given by the cardinally measured index values of $0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 1\frac{1}{4}, 1\frac{1}{2}, 1\frac{3}{4}$, and 2. Under a cardinal index, a car with a value of 1 has twice the quality of a car with an index of $\frac{1}{2}$. The distribution of these cars is shown in Figure 10.1, where the horizontal axis shows the quality level and the vertical axis shows the uniform probability, in this case, of randomly picking a car of each given quality.

Figure 10.1 The Availability of Products of Different Quality (Uniform Probability of Picking Each Car)

Suppose further that a car owner (and potential seller) knows its quality level exactly but that the potential buyers know only the distribution of quality. It is known that the owners have a reserve value on their cars, so that reserve value to the seller = \$5,000 × Q. That is, the owners would sell their cars only if they could get at least \$5,000 for every unit of car quality. In contrast, the nonowners are more eager for used cars and value them at \$7,500 × Q. To make this experiment a complete market, suppose that an auctioneer is hired to call out market prices; sales take place when the auctioneer finds a price that successfully equates quantity demanded with quantity supplied.

DOES A MARKET EXIST? Consider what would happen under asymmetric information. If the auctioneer calls out an initial price of \$10,000 per car, all owners know it is worthwhile to sell their cars, so all nine cars will be supplied. However, nonowners, knowing only the distribution of quality but not the quality of each individual car, will make a best guess that a given car is of average quality; that, is Q = 1. They would not buy any cars at a price of \$10,000 because they are willing to pay only \$7,500 per unit of quality. They guess that all cars have a quality of 1, for a product of \$7,500 × 1 = \$7,500, which is less than the \$10,000 asked. They would be willing to buy cars only if the price were less than or equal to \$7,500.

So the auctioneer, perhaps trying to accommodate the potential buyers, tries a lower price, say \$7,500. Unfortunately, at this price, the owners of the two best cars will withdraw from the market. Why? The owner of the car with two units of quality is receiving only \$7,500 ÷ 2, or \$3,750 per unit of quality; the owner of the car with $1\frac{3}{4}$ units will act the same way. The withdrawal of the two best cars causes the average quality of the seven remaining cars to fall. With nine cars, the average of the distribution was Q = 1. Now at a price of \$7,500 per car, the best car offered will have a quality level of $\frac{3}{2}$, and the average quality will be $\frac{3}{4}$. Potential buyers would now be willing to pay only \$7,500 per unit of quality $\times \frac{3}{4}$ unit of quality per average car, or \$5,625 for any car. Just as the previous price of \$10,000 per car was too high, the new price of \$7,500 is too high for buyers.

Will an equilibrium price ever be found? Surprisingly, in this example, no equilibrium that satisfies both buyers and sellers will be found. The reader can discern this by trying several successively lower prices. In the end, the cars will not be sold even though nonowners value

the cars considerably more than their current owners. Akerlof saw the problem this way: When potential buyers know only the average quality of used cars, then market prices will tend to be lower than the true value of the top-quality cars. Owners of the top-quality cars will tend to withhold their cars from sale. In a sense, the good cars are driven out of the market by the lemons. Under what has become known as the Lemons Principle, the bad drives out the good until, as in some cases such as this one, no market is left.

IMPERFECT VERSUS ASYMMETRIC INFORMATION To see that the problem is asymmetric rather than imperfect information for both buyers and sellers, consider what would have occurred if information had been symmetric. Suppose that *both* owners and nonowners were uncertain of the quality, that they knew only the average quality of used cars on the market. Again, let the auctioneer start with a price of $10,000 per car. All owners, at their best guess, may presume that their car is of average quality, and that the average will again be $Q = 1$. Thus, at a price of $10,000 per car, all nine cars would be offered for sale. However, the nonowners would be willing to pay, at most, $7,500 based on their guess that a given car is of average quality (that is, $7,500 per unit of quality, multiplied by expected quality of $Q = 1$). Again, suppose the auctioneer tries to accommodate the potential buyers by offering a lower trial price, say $7,500. If the owners have imperfect information rather than better information, they will guess that their cars are of average quality, and thus worth (to them) $5,000 per unit of quality, multiplied by the average quality of 1. So the owners are willing to supply nine cars at a market price of $7,500, and the buyers are willing to purchase them at that price. The market thus exists, and clears (supply equals demand) if the information is symmetric—in this case, equally bad on both sides.

This example is extreme in several respects. The assumption of an auctioneer, the assumption that there is only one price for the used cars, the implicit assumption that the parties are not influenced by risk, and even the assumption that the quality of the cars is exogenously given, could each be modified to add more realism. Since the lemons example was published, several analysts have worked on models that modify these assumptions. In some cases, this changes the result significantly. However, Akerlof's main point remains illuminating.

Application of the Lemons Principle: Health Insurance

Adverse selection applies to markets involving health insurance and to analyses of the relative merits of alternative health care provider arrangements. We can apply the Lemons Principle directly to health insurance with the help of the previous example, a mirror image of the insurance problem. In Figure 10.2, let the horizontal axis measure the expected health expenditure levels of a population of n potentially insured people, instead of measuring the quality of used cars. Assume that they have the same demographic characteristics and that their expected health expenditure levels for the insured period range from a low of $0 up to an expenditure level of $\$M$. The vertical axis represents the probability with a uniform distribution (so that the probability of any level of spending is $1/n$). The insurer must at least break even, which means that the premium (or price) received from each buyer must cover the insured population's average expenditure and other expenses (including marketing and overhead).

Information asymmetry will likely occur because the potential insureds know more about their expected health expenditures in the coming period than does the insurance company. To illustrate, assume a potential insured knows his or her future expenditure

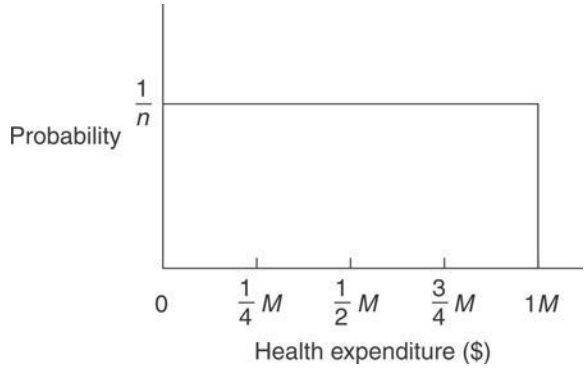


Figure 10.2 Uniform Probability of Expenditure (Expected Health Expenditure Levels)

exactly but that the insurance company knows only the distribution of expenditures for all insured persons.

Again, use the device of the auctioneer to illustrate the point. Suppose this time that the auctioneer attempts a first trial price of $0! All potential beneficiaries would certainly demand coverage at this price. Just as certainly, the insurance company, expecting an average expenditure of $\$\frac{1}{2}M$, would require a premium of at least $\$\frac{1}{2}M$.

Following Akerlof's analysis, suppose the auctioneer tries a higher price, say $\$\frac{1}{2}M$, hoping that this will clear the market. In this case, all potential beneficiaries who expect an expenditure level below $\$\frac{1}{2}M$ will choose to self-insure, that is, leave the insurance market altogether, because this premium is higher than their privately known levels of health expenditure. When these healthier people leave the market, the average expected expenditure level of the remaining insured persons, those with expected expenditures from $\$\frac{1}{2}M$ to $\$M$, rises to $\$\frac{3}{4}M$. Thus, the higher health risks tend to drive out the lower health risk people, and a functioning market may fail to appear at all for some otherwise-insurable health care risks.

Observe again that it is the asymmetry of information rather than the problem of incomplete information that leads to this result. If patients were no better at predicting their health expenditures in our example than the insurer, adverse selection would not take place. That is, all potential beneficiaries would have expected expenditures of $\$\frac{1}{2}M$ and would be willing to purchase insurance at the premium of $\$\frac{1}{2}M$.

Inefficiencies of Adverse Selection

This example illustrates the effects of adverse selection. Health insurance industry analysts recognize that even in its less extreme forms, adverse selection will appear. Even if functioning health insurance markets do evolve in the presence of information asymmetry of this kind, the resulting adverse selection leads to economic inefficiencies.

What are the inefficiencies? Unlike the example, few people can know exactly their future level of expenditures. Risk is the main reason for insurance. However, if the lower risks are grouped with higher risks and all pay the same premium, the lower risks face an unfavorable

rate and will tend to underinsure. They sustain a welfare loss by not being able to purchase insurance at rates appropriate to their risk. Conversely, the higher risks will face a favorable premium and therefore overinsure; that is, they will insure against risks that they would not otherwise insure against. This, too, is inefficient. In addition to inefficiency, income will be redistributed from consumers who are lower risks to those who are higher risks.

Is adverse selection merely a theoretical prediction or a serious problem? Evidence of adverse selection has been found in markets for supplemental Medicare insurance (Wolfe and Goddeeris, 1991) and individual (nongroup) insurance (Browne and Doerpinghaus, 1993). Elsewhere, Cardon and Hendel (2001) found that those who were insured spent about 50 percent more on health care than the uninsured. Although this gap appears to support the existence of adverse selection, it could also be due to two other reasons. First, lowering the price of health care to the patient encourages the consumption of additional health care, that is, moral hazard occurs. Second, the insured may have different observable characteristics, such as age, that are associated with higher spending. However, insurers could incorporate observable characteristics in setting premiums. It is the unobservable characteristics that are the source of asymmetric information, and hence adverse selection. Cardon and Hendel found only a small and statistically insignificant effect of unobservables in explaining the spending gap.

As this study suggests, if information asymmetry threatens to lead to inefficiency and even to the elimination of functioning markets in some cases, we would expect consumers, providers, and insurers to resort to other economic devices and institutions to help mitigate the problem. To illustrate, while the lemons problem in used-car markets is real, a buyer may hire a mechanic to examine the car of interest, the seller may offer a warranty, and agencies or consumer unions may arise to provide quality information. In health insurance markets, beneficiaries were often not covered for pre-existing conditions (including pregnancies) or were charged higher premiums based on health status indicators. These insurance practices, however, are changing with the passage of the ACA.

The Affordable Care Act and Adverse Selection

A prominent and timely example of adverse selection involves the Affordable Care Act, or ACA. A central feature of the ACA is the *individual mandate* which requires most individuals to have insurance or face financial penalties (see Chapter 22). As of March 31, 2014, those without employer-provided insurance or who were not covered by various government programs (such as Medicaid) were required to purchase insurance or face tax penalties. To be eligible for federal subsidies, the insurance must be purchased in what are called *health insurance exchanges*. Individual states can establish exchanges or a state can default to the federal exchange.

Insurance companies that participate in the exchanges faced a daunting challenge as the exchanges were rolled out in October 2013. They may vary premiums based on level of coverage (plans are classified as platinum, gold, silver, and bronze), age, individual or family coverage, geographic location, and tobacco use. Insurers may not engage in the practice of *underwriting*, i.e., charging more or denying coverage for higher risks or those with pre-existing conditions. Renewal is also guaranteed. Insurers that do not accurately predict the mix of patients who would enroll in their plans face serious financial consequences.

Many critics of the ACA are concerned that a much larger number of healthy, young adults would not enroll because of "sticker shock"—premiums that are much higher than those they were previously paying or would have paid had they purchased insurance. It is essential that insurers enroll these low-risk enrollees to help subsidize premiums for sicker

and usually older enrollees and to prevent a vicious cycle of ever-rising premiums. One media source reported that average annual premiums for those aged 18–34 would be about $1,700 but that the benefits would amount to only $350.[2] Such reports imply that it would be foolish for healthy young adults to purchase insurance and that paying the penalty would be their smart financial choice. As of late 2013, younger adults were signing up at less than 60 percent of the rate that was targeted by federal officials and some commentators even predicted the demise of the ACA through the potential escalation of premiums.

Casual observers of the rollout might wonder if those who designed the exchanges anticipated adverse selection, but nothing could be farther from the truth. Insurance experts are well aware of various forms of adverse selection that could arise under the ACA, including switching among plans and between plans offered through exchanges and those offered outside exchanges. As a result, numerous features to mitigate adverse selection were adopted. Examples include: (1) limiting the enrollment period;[3] and (2) the requirement that all plans include an "essential health benefits package."

Risk adjustment is also one of several more sophisticated mechanisms. It rests on a model developed by the Department of Health and Human Services to measure an enrollee's risk score. The risk-adjustment program redistributes funds from plans that attract a disproportionate share of healthy enrollees to plans with less healthy enrollees.

Analysts have yet to determine the effectiveness of features designed to mitigate the effects of adverse selection. To compound matters, with the flawed rollout of the exchanges in late 2013, the essential benefits requirement was temporarily waived under a hardship exemption for some who had their limited "catastrophic" plans cancelled. Most of those seeking the waiver were likely to be young and healthy, compounding the transition to the new health insurance marketplace. Numerous waivers and rule changes that have been granted since the ACA was approved in 2010 weaken the safeguards against adverse selection. These observations leave little doubt that adverse selection will present ongoing challenges to both the insurance industry and ACA administrators.

Experience Rating and Adverse Selection

Group insurance can often be a useful mechanism to reduce adverse selection. Most employees and their families in the United States are insured through employer group plans rather than through individual policies. Group plans enable insurers to implement experience rating, a practice where premiums are based on the past experience of the group, or other risk-rating systems to project expenditures. Because employees usually have limited choices both within and among plans, they cannot fully capitalize on their information advantage.

Although experience rating can reduce adverse selection, it has come under increasing attack with the rapid growth of managed care plans such as health maintenance organizations (HMOs). HMOs receive predetermined premiums to provide the contracted health care for their enrollees. In contrast to traditional insurance where providers are independent of insurers, HMOs integrate insurance with the provision of health care.

Policymakers have promoted HMOs and other managed care organizations heavily in the belief that they have a powerful self-interest in eliminating the inappropriate care that might be recommended by providers in traditional fee-for-service systems. However, the intense competition to enroll healthy populations in managed care plans, known as "cream skimming" or "cherry picking," has led to concerns that insurers are more interested in finding favorable groups than providing quality care. Experience rating also redistributes income toward healthy populations, and the perceived inequity of such redistributions has become a public issue.

To deal with these concerns, several states have introduced some degree of mandated community rating—a practice in which an insurer charges all groups within an area the same premium. In a less rigid form, upper and lower limits on premiums are established through rate bands. The effects of these changes have not yet been determined, but Goldman et al. (1997) predict some serious redistributional consequences of the community rating schemes considered for California. In particular, because health care spending in wealthy, urban areas is relatively high, their model predicts large regional transfers of income to the urban areas from poorer, rural communities.

The Agency Relationship

An agency relationship is formed whenever a *principal* (for example, a patient) delegates decision-making authority to another party, the *agent*. In the physician–patient relationship, the patient (principal) delegates authority to the physician (agent), who in many cases also will be the provider of the recommended services. The motive behind this delegation of authority is that the principals recognize that they are relatively uninformed about the most appropriate decisions to be made and that the deficiency is best resolved by having an informed agent. Thus, asymmetric information and agency are closely related phenomena.

Agency and Health Care

What would the perfect agent do? The perfect agent physician is one who chooses as the patients themselves would choose if only the patients possessed the information that the physician does. (See Box 10.1 for a study in which physicians and other "experts" are the patients.) This is in line with the medical code of ethics to the extent that the patient's own interest focuses on his or her health. When conflicts arise, perfect agents focus on the patients' preferences, not their own.

The problem for the principal is to determine and ensure that the agent is acting in the principal's best interests. Unfortunately, the interests may diverge, and it may be difficult to introduce arrangements or contracts that eliminate conflicts of interest.

As an example, Dranove and White (1987) ask why we do not reimburse physicians on the basis of improvements in patient health. More simply, why are they not reimbursed only if the patient is cured? It would appear that such a contract would arise naturally to merge the interests of both principal and patient. The authors suggest that such contracts do not exist because of the problem of asymmetric information, although in this case it is the physician who may lack information about the patient's well-being.

BOX 10.1

What Happens When the Patient Is a Medical Expert?

In the United States and some other countries, observers have raised concerns over the high and increasing rates of caesarean sections (C-sections). Some have attributed this phenomenon in part to imperfect agency where the financial self-interests of providers lead to them to recommend and perform more C-sections than would otherwise take place. They have suggested that the higher rates of reimbursement for privately

insured than publicly insured patients in the United States act as a principal determinant of the substantially higher C-section rates for the former group (though Grant (2009) estimated that other factors account for much of the gap).

What happens, however, when the expectant mothers are themselves experts: physicians, obstetricians, or midwives? Grytten and colleagues (2011) examined childbirths in Norway where obstetricians receive fixed salaries and where hospitals are under tight budget controls so that unnecessary C-sections would be discouraged. The authors sought to understand the role of education, hypothesizing that less educated mothers would be the easiest to persuade to have a normal vaginal delivery when complications arise, and that expert mothers would be the most difficult to persuade.

Agency theory predicts that expert mothers would have higher C-section rates, which is the case in Norway, but this prediction is also consistent with an alternative known as "statistical discrimination." Under the statistical discrimination model, the preferences of expert patients are more closely met because they are bettor communicators. Grytten's research methodology distinguishes between these competing hypotheses by considering new technologies that help detect fetal distress and consequently reduce clinical uncertainties regarding the appropriate delivery. Under the statistical discrimination hypothesis, these new technologies should reduce disparities in C-section rates between expert and non-expert mothers. The disparities would not decrease under the agency model because the new technologies enable providers to retain their information advantages and influence over non-expert mothers. Empirical estimates show decreasing differences between expert and non-expert mothers over 1967–2005, supporting the statistical discrimination theory. In other words, the new techniques did reduce disparities.

To illustrate, consider patients with low back pain. Regardless of their improvement, the patients have financial incentives to understate the extent of their improvement. The provider also has an incentive to overstate the difficulty in treating the patients and in improving their health in order to increase the payment (which, let us assume, is based on the difficulty of the case). Further, it is these information problems and not other special characteristics of health care delivery that preclude payment based on the degree of improvement of the patients' conditions.

Dranove and White further apply agency theory to explain other features in the organization of health care delivery. Patients often establish a long-term relationship with a physician and pay that physician on a fee-for-service basis. As discussed earlier, such an arrangement would appear prone to lead to conflict between patient and provider. It is thus natural to ask the following questions: Why does this particular physician–patient arrangement arise and why is it so common?

Dranove and White argue that a continuous relationship between patient and physician provides the patient with increasing information with which to monitor the physician. This information places constraints on the extent to which the provider is able to deviate from an agency responsibility. Monitoring also encourages the physician to make appropriate referrals to other providers when he or she is unable to provide the services alone.

We can add to their argument by pointing out that a continuous relationship reduces the cost of transferring information about medical history, circumstances, and preferences from patient to provider. These advantages of the usual physician–patient relationship would be eroded if patients and providers were to switch to limited-period contracts under which providers are reimbursed on a different basis.

Consumer Information, Prices, and Quality

We next examine the effects of imperfect information on the price and quality of medical services. Would relatively poor consumer information reduce the competitiveness of markets? Does increasing physician availability increase competition and lower prices as traditional economics suggests? What happens to quality? How do consumers obtain and use information? Several studies provide helpful insight.

Consumer Information and Prices

Satterthwaite (1979) and Pauly and Satterthwaite (1981) introduced one of the most novel approaches to handle issues involving consumer information and competition. The authors identify primary medical care as a reputation good—a good for which consumers rely on the information provided by friends, neighbors, and others to select from the various services available in the market. Physicians are not identical and do not offer identical services. Because of this product differentiation, the market can be characterized as monopolistically competitive.

REPUTATION GOODS Under these conditions, the authors show that an increase in the number of providers can increase prices. The reasoning behind this surprising prediction is logical. Recall that a typical consumer relies on other consumers for information regarding their experiences with physicians. Thus, when physicians become numerous, the average number of friends who see any provider diminishes; this, in turn, diminishes the average level of information available. The consumer's responsiveness to prices and other practice characteristics depends on his or her knowledge of—that is, information about—the available alternatives. Thus, this reduced information reduces the price responsiveness (i.e., the elasticity) of the firm demand curves, causing the equilibrium prices to rise. The economic idea is that reduced information tends to give each firm some additional monopoly power.

Reduced information enhances monopoly power and reduces the elasticity of the firm demand curves; this is consistent with standard theory. That such a situation may arise as the number of sellers increases is an unconventional and counterintuitive idea. The authors have, however, provided empirical support for their theory, and the interested reader is referred to their work for further study.

THE ROLE OF INFORMED BUYERS The degree to which imperfect price information contributes to monopoly power should not be overemphasized. Recall that it is not necessary for every buyer of a commodity to have perfect price information to elicit relatively competitive pricing conditions. Realistically, most consumers lack complete price information about many of the goods and services they buy (they don't know what alternative sellers are charging). Yet, despite variations in the prices of individual items among, for example, grocery stores, the average charges for a set of items across similar types of stores are likely to be similar and close to competitive pricing. A growing body of literature shows that it is sufficient to have enough buyers who are sensitive to price differentials to exert discipline over the marketplace. This will likely hold especially where the damaging threat exists of having any systematic differentials publicized by consumer organizations or the low-priced merchants themselves.

These arguments suggest that while imperfect price information will likely produce higher prices, this phenomenon may be substantially limited. In health care markets, where many services are fully or partially covered by insurance, there are added considerations. While a patient may become less sensitive to price levels and price differentials in the choice of providers, third-party payers, such as insurers, have assumed a monitoring function. Through

selective contracting and other fee agreements, the actual reimbursement is often lower than the provider's charges.

PRICE DISPERSION The distinction between the effective transaction price and a provider's charge also obscures evidence of dispersion of fees as distinct from the average level of fees. Under conditions of imperfect consumer information, Nobel Laureate George Stigler (1961) argued that variation in prices will increase.

Building on Stigler's insight, Gaynor and Polachek (1994) developed measures of the degree of both buyer and provider ignorance by using frontier regression methods. These authors separated price dispersion into measures of incomplete buyer information, incomplete seller information, and random noise. They found that both patients and physicians exhibited incomplete information with the measure of ignorance being one and one-half times larger for patients than for physicians.

Consumer Information and Quality

Many reports have documented high rates of medical errors and inappropriate care. For example, McGlynn and colleagues (2003) evaluated the medical records over a two-year period for a random sample of adults in 12 metropolitan areas. The study participants received only 55 percent of the recommended care overall with about the same proportions for recommended preventive care (55 percent), recommended acute care (54 percent), and care recommended for chronic conditions (56 percent).

Because consumers cannot easily monitor quality, the search for information regarding quality can be costly. At the same time, the consequences of poor-quality care can be severe or even fatal. Thus, as the previous discussion suggested, despite asymmetric information, patients rely on a variety of countervailing arrangements that are intended to reduce their search costs. These include licensure and certification, the threat of malpractice suits, codes of ethics, and various quality-assurance schemes that are either mandated or voluntary. The Internet is also becoming a major source of information despite concerns about the accuracy of online information.

Patients' abilities to make rational choices including those that involve quality attributes are not easily observed. These choices are especially important to strategies that involve trade-offs between regulatory and market-based reforms. Although some research (Waber et al., 2008) raises concerns about the ease with which consumer quality perceptions can be manipulated, two "real world" economic analyses give comfort to the proposition that consumers respond to improved quality in expected ways.

Leonard (2008) examined whether patients in Tanzania (presumably less sophisticated than those in United States and other developed countries) can detect changes in the quality of care provided by physicians, as measured by physicians' adherence to treatment protocols. The study design takes advantage of the Hawthorne effect, which describes a temporary change to behavior in response to a change in the environmental conditions—in this case, significantly improved provider adherence through observation by a research team. Adherence slowly returned to usual levels even though the physicians continued to be observed. Leonard examined patient responses to the temporary increases in quality and found that every 1 percent increase in adherence to protocols increased the probability that a patient will be very satisfied with the doctor's quality by 0.4 percent. From the responses to the equivalent of an experimental change in quality, he concluded that patients "recognize and value quality care."

Howard (2005) examined registrations of candidates who were suitable for kidney transplants. These transplants are typically performed at major medical centers, and many factors

in addition to perceived quality can influence patient decisions, for example, provider recommendation, distance to the medical center, insurance restrictions, and quality. Using graft failure as a measure of quality, Howard found that a one standard deviation increase in the one-year graft failure rate, a clear indicator of poor quality, was associated with a 6 percent reduction in patient registrations at a center.

At this point, we pursue the consumer's direct role through the Dranove and White argument that the physician–patient relationship enables patients to monitor providers and encourages physicians to make appropriate referrals. To the extent that many specialists rely on referred patients, these specialists would seem to have incentives to maintain quality. Are they also rewarded with higher prices for higher-quality services? Theory suggests that if consumers have the ability to distinguish between quality levels, then the demand for higher-quality providers and thus price should be greater than for lower-quality providers. Haas-Wilson (1990) examined this proposition using data from the psychotherapy services market. She investigated whether the prevalence of referrals from informed sources affects the price of social workers' psychotherapy services. Informed sources include other health providers and other professionals such as school counselors and clergy.

Regression analysis of a sample of social workers' fees indicated that fees are positively and significantly affected by the percentage of clients who were obtained through informed referrals. The evidence shows that patients rely on informed sources (agents) for information and that higher quality, as measured by informed referrals, is rewarded by higher fees. This evidence, however, should not be taken to mean that consumers necessarily associate higher fees with higher quality. In a recent U.S. survey conducted by Phillips and colleagues (2016), a majority of the respondents did not associate quality with health care prices.

Other mechanisms can also help reduce the problems created by asymmetric information. As previously discussed, the lack of a clear profit motive may make nonprofit organizations more attractive to patients when they cannot easily observe or determine quality. Although this argument seems plausible, empirical support for it had been lacking until Chou (2002) developed a novel application to quality of care in nursing homes (where nearly two-thirds of the homes are for-profit). Because many nursing home patients are too cognitively or physically incapacitated to be able to monitor and evaluate their care, Chou used the absence of visits by a spouse or children (the patient's representatives) within one month of admission as an indicator of information asymmetry. There were no significant differences between for-profits and nonprofits when asymmetric information was not present. With asymmetric information, a very different picture emerged. For-profits had higher mortality rates as well as higher rates of decubitus ulcers, dehydration, and urinary tract infection. Chou concluded that for-profits "have more incentive to compromise on those aspects of quality of care which are hard to monitor."

Other Quality Indicators

With the increasing dominance of HMOs and other managed care organizations, the availability of information to help consumers select among plans, and to monitor how consumers respond to the information, have become major issues. The National Committee for Quality Assurance (NCQA), a private accreditation body for HMOs, issues report cards based on about 50 standardized measures of a plan's performance (such as childhood immunization rates, breast cancer screening, and asthma inpatient admission rates). *Newsweek*, *U.S. News & World Report*, and various consumer groups also regularly rate HMOs. (See Box 10.2 for recent evidence on the relationship between patient outcomes and the *U.S. News & World Report* hospital rankings.) A key assumption behind these efforts is that information

about quality will, like price information, help discipline providers through patient choices. Low-quality HMOs will presumably not survive, or at least they will not be able to charge high-quality prices.

BOX 10.2

Quality Rankings and Health Care Outcomes

The annual *U.S. News & World Report* national hospital rankings are eagerly anticipated by the medical community and the public. How do these quality rankings stack up against patient outcomes? White and colleagues (2014) examined relationships between hospital prices, various quality indicators and other hospital characteristics. In addition to the *U.S. News & World Report* rankings, based on reputation among physicians, they used outcomes measures developed by the Centers for Medicare and Medicaid Services (CMS).

The high-price hospitals were, on average, much larger than low-price hospitals. They also had much larger market shares. Not surprisingly, the high-price hospitals outperformed low-price hospitals on *U.S. News & World Report* ratings (no low-price hospital was on the list). However, high-price hospitals generally performed the same or worse on most objective quality indicators e.g., postsurgical death rates and serious blood clots among surgical discharges. If these findings are confirmed through more extensive research, they would clearly have profound implications for everyone involved in selecting or referring patients to hospitals and for third-party payments to hospitals.

Initial evidence on the intended effects of plan performance ratings brings the report card strategy into question. Tumlinson et al. (1997) found that independent plan ratings are relatively unimportant to consumer choices. Only 17 percent of survey respondents indicated that such ratings are essential, compared with 72 percent for specific plan benefits and 62 percent for out-of-pocket costs. Chernew and Scanlon (1998), employing multivariate statistical methods on consumer choice of plans, confirm that "employees do not appear to respond strongly to plan performance measures, even when the labeling and dissemination were intended to facilitate their use" (p. 19).

In subsequent work, Scanlon and colleagues (2002) examined a flexible benefits system introduced by General Motors in 1996 and 1997 under which employees and retirees received a fixed amount of dollars that could be spent across a variety of fringe benefits categories. Excesses in spending over the allotted "flex dollars" are paid out-of-pocket. GM developed ratings (e.g., superior performance, below expected performance) for six performance categories (e.g., preventive care, access, patient satisfaction) to help those who wanted to select an HMO choose among the available plans (typically two to six depending on the employees' geographic area). As expected, the study found that higher out-of-pocket prices imply lower enrollments. Also as expected, employees tended to avoid plans with many below-average

ratings. However, plans with many above-average ratings were not much more successful in attracting enrollees relative to plans with many average ratings.

More recent contributions for very different medical interventions also provide new evidence on report cards and how they affect patient decisions. Wang and colleagues (2011) examined coronary artery bypass graft surgery (CABG) in Pennsylvania, a state that has published report cards on CABG providers since 1992. Beginning in 1998, the report card information for both hospitals and individual surgeons (e.g., number of cases, mortality, and readmission rates) were made available on the Pennsylvania Health Care Cost Containment Council Web page (www.phc4.org/). The Wang study considers both demand side effects (e.g., decrease in demand for poor performance surgeons) and supply-side effects (e.g., "dumping" or avoiding high risk patients) of report card information. At the surgeon level, volume decreased for poor or unrated surgeons indicating a patient avoidance effect. Surprisingly, though there was no increase for high performing surgeons, and also no effects on hospital volume of any rating.

Unlike most CABG procedures, fertility treatment involves completely elective procedures that are not often covered by insurance. Does information on fertility clinics' success rates affect clinic choice? You bet it does! Bundorf et al. (2009) compare the three-year lagged birth rates of fertility clinics before and after report card information became publicly available in 1995. The authors found that the differential effect of the pre- and post-reporting for clinics with higher birth rates increased, thus supporting the hypothesis that consumers respond to quality information.

Conclusions

There is little doubt that information gaps, asymmetric information, and agency problems are prevalent in provider–patient transactions. However, for some health care services, the problems are not necessarily greater or larger than those for other goods. Patients are likely to be relatively poorly informed about treatment for conditions that they have not previously experienced and about care involving newer technologies. The informational asymmetries and reliance on provider-agents are likely to be most pronounced in these situations.

Although there is a potential lack of competition, even wide information gaps do not necessarily lead to market failure. Leaving aside the role of licensure and regulation, arrangements have evolved to help patients or their insurers to monitor the quality and prices of providers. Furthermore, higher-quality producers are generally rewarded by greater demand and higher prices. The use of referrals, accreditation, and other arrangements reduce the provider's ability to raise prices above those charged by others and to sell low-quality services at high-quality prices. Nevertheless, improved quality remains an elusive national goal. Despite efforts to provide quality information and many private and government initiatives to improve quality, a series of influential reports released by the Institute of Medicine (1999, 2001) suggested that as many as 98,000 deaths and a million excess injuries annually in the United States can be attributed to problems with quality and safety. By calling for a fundamental overhaul of the U.S. health care system, the Institute raised troubling questions about existing safeguards as well as patients' perceptions of quality and their ability to monitor it.

In the years since those publications, the Agency for Healthcare Research and Quality (AHRQ), which evaluates quality through an annual report covering numerous measures of quality, safety, and effectiveness, found that while quality is improving, the rate of

improvement varied by measure with 40 percent of the measures showing no improvement at all (AHRQ, 2015). Its conclusions echo an earlier assessment by the Commonwealth Fund (2008) which warned that the "U.S. health system is on the wrong track." From a National Scorecard of 37 indicators of health system performance based on comparisons with top performing states, regions, health plans, and other nations, the Fund concluded (p. 13) that the United States "is losing ground in providing access to care and has uneven health care quality. Average U.S. performance would have to improve by 50 percent across multiple indicators to reach benchmark levels of performance." The challenge is to develop and implement system-wide reforms that can narrow the gap between processes that work and what is actually done.

To meet this challenge, the ACA has established quality improvement reporting requirements for most private insurance plans. Emphasis will be placed on improvements in health outcomes and patient safety; health promotion and wellness; and on reductions in medical errors and hospital readmissions. Among other measures, CMS was directed to test new methods of delivering care. It will also provide financial incentives for the formation of integrated entities such as "Accountable Care Organizations" that are held responsible for improving quality while holding down costs. As with many features of the reform act, the effectiveness of its quality improvement measures has yet to be evaluated.

The ACA has also renewed interest on adverse selection including the challenge of distinguishing between risk selection and moral hazard (Bajari et al., 2014). We have described some of the features of the ACA designed to mitigate adverse selection. The rapid growth of consumer-directed and other high-deductible health plans (see Chapter 22) is another recent phenomenon that could lead to significant risk selection. One would expect these plans to attract younger and healthier enrollees leaving high cost enrollees in more traditional plans. McDevitt et al. (2014) did indeed find such evidence. But they also found that adverse selection could be reduced by appropriate financial and plan attribute features. These include employer account contributions and requiring employees to make active plan choices rather than being placed in their current or default plans.

Summary

1 Health care markets tend to be characterized by both imperfect information and asymmetric information. Asymmetric information describes a situation in which those on one side of a transaction have better information than those on the other side.
2 Often, providers are relatively well informed (e.g., about the patient's illness and possible treatments). In other cases, buyers are relatively well informed (e.g., the purchaser of insurance knows more about his or her health status and pertinent habits than the insurer does).
3 The extent of consumer information problems should not be exaggerated. Consumers are reasonably well informed on about one-fourth to one-third of their health care spending.
4 One possible consequence of asymmetric information is that a market will not exist. Even if it exists, a general reduction may occur in the quality of goods available (the "Lemons Principle").
5 The Lemons Principle appears as the problem of adverse selection in health insurance and other health care markets.
6 Adverse selection results from asymmetric information, not equally imperfect information. Adverse selection in insurance results in inefficiencies through higher-risk consumers

overinsuring, relative to the amounts they would purchase at actuarially fair rates, and lower risks correspondingly underinsuring.

7 The ACA introduced various features under the health insurance exchanges to limit adverse selection. It is yet unclear how well these features will work.

8 An agency relationship tends to be formed when a party (principal) delegates decision making to another party (agent). The problem for the principal is to develop a contract or relationship to ensure that the agent is acting in the principal's best interests.

9 Various agency relationships have evolved to mitigate the problems associated with asymmetric information between patient and provider. These include the continuous physician–patient relationship and the health maintenance organization.

10 Other constraints, such as licensure and accreditation, codes of ethics, and the threat of litigation, limit the ability of providers to deviate from their agency responsibilities.

11 Many health care services are reputation goods. In markets for reputation goods, an increase in the number of providers can lead to an increase in monopoly power and higher prices.

12 The existence of informed buyers helps exert discipline over the market by limiting price increases and price differentials among sellers.

13 Though challenged by some, the proposition that higher quality tends to be rewarded with higher price is supported in economic studies. Patients also respond to quality indicators in selecting a hospital. However, they rely only modestly on objective plan ratings in their selection of HMOs. Negative ratings have a greater impact on consumer decisions than positive ratings.

14 Improving quality has become a national priority. Many studies have found high levels of inappropriate care and medical errors. The ACA contains various provisions that are designed to improve a variety of quality indicators.

15 The ACA as well as the growth of high-deductible health plans are creating opportunities for adverse selection. Health economists are trying to estimate the extent of this phenomenon and the effects of various tools that could mitigate it.

Discussion Questions

1 The market for higher education is another example where a high degree of information asymmetry is likely. What mechanisms have evolved to help students in their choice of schools and classes within schools? Do you have confidence that higher-priced institutions provide higher-quality education?

2 The situation in which an individual is interviewing for a job also exhibits information asymmetry. Explain why. How does the relatively poorly informed party deal with this?

3 The use of professional and independent buyer-agents to help individuals purchase automobiles or houses is becoming a more common phenomenon. Given the conflict of interest facing the physician-agent, why do we not see greater use of a buyer-agent who is retained by the patient?

4 The used-car market has publications that provide information on the quality and prices of used cars. Are similar avenues of information available to health consumers? What kind of information do they provide? Is it more or less effective than the information available on used cars? How would you, as a patient, find information about a provider's quality or prices? How would you assess the confidence you have in that information?

5 What is a reputation good? What are examples of reputation goods outside the health care sector? Show what Pauly and Satterthwaite predict will happen to the demand curve for health services as a result of an increase in the number of providers.
6 Stigler argued that the variation in fees increases as buyer information decreases. Suppose you observe that each seller in a market is charging the identical price. What potentially conflicting inferences can you draw?
7 Why don't physicians guarantee their work as do many auto repair shops?
8 Various commentators have suggested that only 15 to 20 percent of all health care services have been subject to rigorous, controlled investigation, that is, care based on what is commonly called "evidence-based medicine." Assume that this statement is correct. What are some implications for efficiency of health care delivery?
9 Is it possible to have a situation where higher costs, as measured by the resources used to provide care, do not produce higher quality?
10 According to clinical research, nearly one-half of the care provided in the United States falls short of recommended treatment protocols. Discuss how imperfect and asymmetric information contribute to this phenomenon. How can health plans or markets be reformed to reduce the quality gap?
11 According to some analysts, the success of the ACA is dependent on having most young and healthy adults comply with the individual mandate to buy insurance. What will happen if young adults enroll at unexpectedly low rates?
12 Describe several features of the ACA that are designed to minimize adverse selection. Why might some of those measures not end up working very well?

Exercises

1 Suppose that in the Akerlof example, there are only eight cars ranging in quality from ¼ to 2 (i.e., there is no complete lemon). Hence, the mean quality level is 1.125. Determine whether the market disappears completely, and, if not, how many cars will be sold.
2 Consider the agency relationship in malpractice cases under a contingency fee system. The plaintiff (party that sues) typically pays his or her attorney about one-third of any monetary damages that are awarded (and nothing if the case is lost). Supporters of this system claim that client and attorney share a common goal of maximizing the award. Is there, however, an inherent conflict between attorney and client in the amount of attorney time and other resources that are devoted to the case?
3 Give three examples of asymmetric information in which the health consumer has information that is unavailable to the health provider. Give three concrete examples in which the health provider has information that is unavailable to the health consumer.
4 In the Akerlof example, the individuals are treated as indifferent to risk. What would you expect to see in these markets if individuals wanted to avoid risk? What if there were some "risk lovers"?

Notes

1 The emergence of health economics as a distinct field is often traced to Kenneth J. Arrow's (1963) seminal article, "Uncertainty and the Welfare Economics of Medical Care." Arrow emphasized the role of imperfect information and uncertainty, especially the features of health care markets due to the "imperfect marketability of information." A special issue of

the *Journal of Health Politics, Policy and Law* (October 2001) examines Arrow's contribution within a contemporary context.

2 Dan Mangan, "Obamacare Math May Not Add up for 'Young Invincibles': Study," cnbc.com/id/01343022, January 18, 2014. Other sources for this section include: Blumberg and Rifkin (2013); Amy Goldstein and Sandhya Somashkehar, "Health-Insurance Sign-ups by Young Adults Are off Pace Seen as Key to New Law's Success," *Washington Post*, January 13, 2014, washingtonpost.com/national/health-science/young-adults-make-up.cnbc.com/, January 18, 2014; and Michael J. Boskin, "ObamaCare's Troubles Are Only Beginning," *Wall Street Journal*, December 15, 2013.

3 As of January 2016, there have been over 30 "special enrollment" periods under the ACA. These can create powerful incentives for adverse selection. A representative for the Blue-Cross Blue-Shield Association interviewed by the *New York Times* noted that enrollees signing up in special enrollment periods have 55 percent higher utilization than those signing up during the regular enrollment period. See Robert Pear, "Insurers Say Costs Are Climbing as More Enroll Past Health Act Deadline," *New York Times*, January 9, 2016, http:nyti.ms/1OYMuCM.

Chapter 11

The Organization of Health Insurance Markets

In this chapter

- Loading Costs and the Behavior of Insurance Firms
- Employer Provision of Health Insurance: Who Pays?
- Employer-Based Health Insurance and Labor Supply
- The Market for Insurance
- The Uninsured: An Analytical Framework
- Impacts of the Affordable Care Act on the Uninsured
- Conclusions

Chapter 8 introduced the concept of insurance, an arrangement that allows risk-averse people to reduce or eliminate the risks they face, with a primary focus on health insurance. Consumers buy insurance to replace the uncertainty of a large loss or major expenditure with the more certain prospect of regular premiums. In most countries, profit-seeking firms supply various types of insurance, although the provision of health insurance varies from country to country. In an idealized market, the insurance premium (as a percentage of the potential loss) will approach the probability of the event occurring.

In previous chapters, we concentrated on the impact of insurance on individuals. In this chapter, we focus on the insurance market and the behaviors of firms within that market. Within the context of the employer-provided health insurance common in the United States, we establish who pays for health insurance. We continue with an examination of employer-provided insurance and job mobility. We then look at the traditional community-rated health insurance (where individuals or groups all pay the same premium) and show how that market has changed, and we follow with an analysis of the uninsured. We finish the chapter by examining the impact of the Affordable Care Act (ACA) and its mandated coverage on the level and percentage of those who are uninsured.

Loading Costs and the Behavior of Insurance Firms

Consumers can improve their well-being through insurance by sacrificing a (relatively) small but certain premium to insure against the probability of a considerably larger loss. It is important now to demonstrate how within competitive markets the policies will be offered to specific groups and why, in fact, some groups may find it difficult to get insurance at all.

We have referred to the model of a competitive industry, in which the firms will compete to where economic profits become zero, or normal. With higher (lower) profits, firms will enter (leave) the market. Only when profits are zero, or normal, will entry and exit cease. In this model, the insurance carriers collect money during the year and pay some of it out. In good years, carriers pay out less than collected; in bad years, they pay out more. Economic analysis suggests that the good (and bad) years will be random. Systematically good (bad) years suggest excess profits (losses), and the probability of entry into (exit from) the industry by other firms.

We also have previously shown how moral hazard can lead firms to offer certain types of coverage and not others. In particular, firms have often shown themselves to be reluctant to cover conditions accompanied by price-elastic demands for services.

Impacts of Loading Costs

Insurance firms incur costs of doing business that are added to the claims payouts. These loading costs are largely related to the numbers and types of customers and claims processed. Even in perfect competition, these costs must be passed on to consumers, or else the insurers will not be able to cover all costs and will be forced to leave the market. The incidence of these costs suggests that insurers will shy away from covering events that are almost certain to occur, or those that seldom occur.

Consider consumers who behave as though they have a utility of wealth (W) function exhibiting diminishing marginal utility of wealth. Figure 11.1A relates total utility to total wealth and Figure 11.1B looks at corresponding marginal gains and marginal costs related to various actions. In Figure 11.1A, Sara has $20,000 in wealth yielding utility at point A, with various possibilities of losses up to $10,000, or point B. The amount Sara would be willing to pay over the actuarially fair amount (also interpreted as Sara's consumer surplus) is shown by the horizontal distance between the expected utility line and the (curved) utility function, measured in dollars. For example, at point F, this horizontal distance is FG. On inspection we note that the horizontal distance between the expected utility line and the utility function is zero if the event never occurs (i.e., if we are at point A). It increases up to some point as we move in a southwest direction (with increased probability of illness) and then decreases to zero, as the illness becomes more certain, toward point B.

Because insurance is taken against risk, as the probability of the uncertain event approaches either 0 or 1, insurance becomes less desirable. Near point A, the expected loss—that is, the probability of the event—multiplied by the loss if the event occurs, is not large enough for Sara to bother to insure. This is noted as point A in Figure 11.1B, the lower diagram, where dollars replace units of utility on the vertical axis. Going back

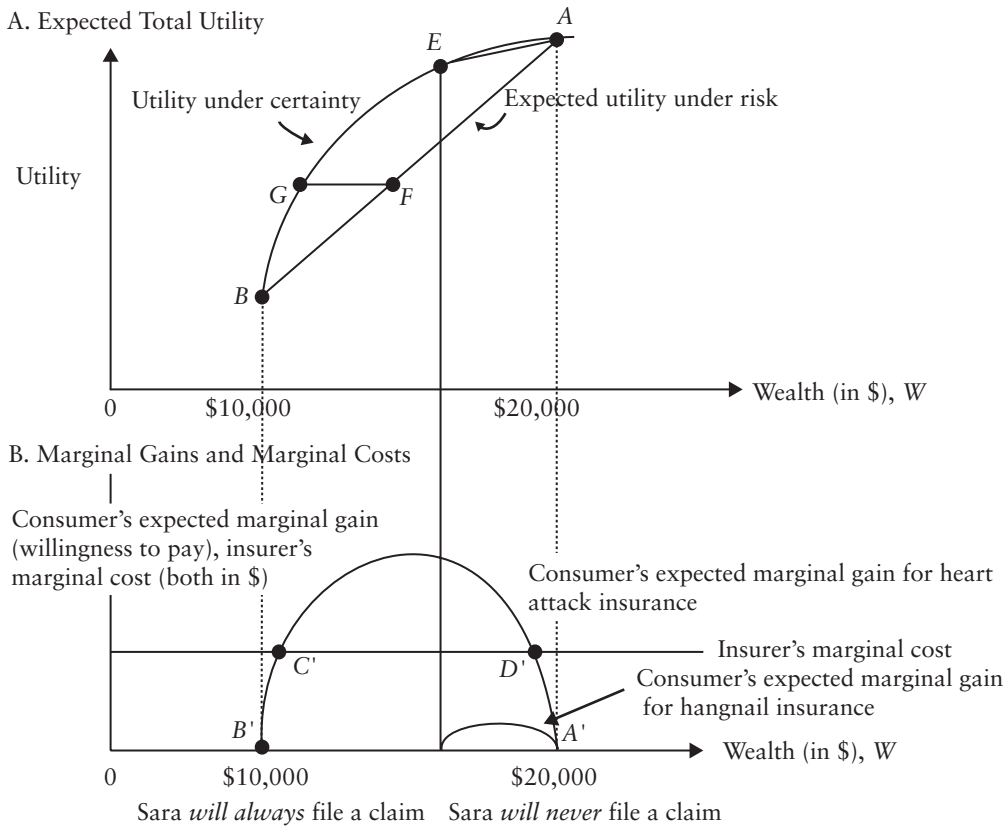


Figure 11.1 Impacts of Loading Costs on Availability of Insurance

to Figure 11.1A, at point *B*—because the event is almost certain—Sara might as well set the money aside (self-insure) and avoid the trouble of dealing with the insurer. The corresponding point on Figure 11.1B is *B′*.

Insurance for Heart Attacks and Hangnails

In comparing types of losses, for any probability of illness, the larger the expected loss, the larger the gain from the insurance. We see this in Figure 11.1A by comparing the distances between the expected utility line and the utility curve for a small loss (line segment *EA*) and for a large loss (line segment *BA*). Segment *EA* shows a small distance; segment *BA*, a larger one. Hence, if Sara has equal probabilities of a hangnail (small loss) and a heart attack (large loss), her expected gain from heart attack coverage will exceed the expected gain for hangnail coverage.

Consider now the insurers' decisions in providing insurance. If the event is almost certain, the insurers' costs of administering the policy may exceed the benefits to the consumers. In Figure 11.1B from *B′* to *C′*, it will not pay to insure claims because the marginal costs they must charge, to earn profits, exceed the expected consumers' marginal benefits. Between points *C′* and *D′* expected marginal benefits exceed marginal costs. To the right of point *D′*, again the marginal costs exceed the expected marginal benefits, and no insurance will be provided. As the diagram is drawn, no firm could afford to offer hangnail coverage.

Loading Costs and the Uninsured

The forthcoming discussion models the mixed system in the United States, in these formative years following the 2010 passage of the Affordable Care Act. The ACA in large part sought to build upon the existing employer-based system by mandating that individuals purchase insurance coverage, largely within the market context described here. We will spell out details of the Affordable Care Act in detail in Chapter 22, but we will look here at specific ACA impacts on the uninsured.

The analysis of loading costs provides one avenue for addressing the problem of those who cannot get insurance. Health insurance in the United States has been largely available through participation in the labor market. Those who do not participate in the labor market, and many of those who are employed by small businesses, self-employed, or sporadically employed, have found it difficult to get insurance.

Many explanations have been proposed, but it is apparent that the per-person costs of processing information and claims of those individuals who are outside larger organizations (either companies or unions) are higher. This results in an increase in the firms' marginal costs relative to the consumer's marginal benefits and can reduce or eliminate the range of services that may be offered.

The analysis also helps address the impacts of entry and exit in the insurance market. More efficient processing and information handling presumably will lower the premiums that must be paid by customers in the market. If we look again at Figure 11.1B, we recognize that improved information handling and processing would not only lead to lower marginal costs and hence lower prices, but also would permit firms to offer services (based on probability of occurrence) that had not previously been offered.

Consider points *C′* or *D′*, where the expected marginal benefit was previously just equal to (or possibly just below) the marginal cost. An insurer who lowers costs can offer coverage for types of events that previously were uncovered. Conversely, increased costs, due either to market forces or to mandated coverage, would force firms to cut back coverage on events

for which they could not (due to limited consumer surplus) pass along the increased costs on to the customers.

Employer Provision of Health Insurance: Who Pays?

The largest segment of the American population acquires health insurance through the workplace, and this began almost by accident in the 1940s. During World War II a booming economy coupled with wartime shortages left consumer goods in short supply, so the federal government imposed wage and price controls as anti-inflationary devices. Predictably, employers had to devise new ways to attract workers because wage controls in a full-employment economy prevented companies from raiding workers from one another. Fringe benefits were not legally considered as part of the wage package, so they could provide flexibility in worker compensation, improving the allocation of workers among sectors of the economy. One of these fringe benefits was health insurance.

Economists start their analyses by looking at the labor market. We assume that a lower market money wage rate leads an employer to hire more workers for two reasons: (1) the employer can substitute labor for more expensive equipment or resources; and (2) the employer can sell more products at lower prices, hence requiring more workers. Assume at the outset there is no health insurance benefit, and that the market wage is $20 per hour. Employers will hire workers as long as the incremental (marginal) revenue from the goods those workers produce exceeds the $20 per hour wage. To begin, assume that the employer employs 1,000 workers, at an equilibrium money wage of $20 per hour.

Suppose that workers negotiate a health insurance benefit worth $1 per hour to them, and costing exactly $1 for the employer to provide. The employer, who was previously willing to pay a wage of $20, will now be willing to pay $20 less the $1 cost of providing the benefit. Other points on the employer's demand schedule, showing the number of workers it would hire at different wages, will also change by the $1 cost of the benefit. This takes care of the demand side.

Workers supplying their labor to the firm would prefer the $1 benefit on top of the previous $20 wage, but if they were previously willing to accept a wage of $20, they will now be willing to supply their labor for $1 less, because they value the benefit at $1. In the resulting equilibrium, the net wage or total compensation (money wage + the value of the benefit) remains unchanged at $20, but the equilibrium money wage falls to $19, or by exactly the amount of the benefit. Workers accept lower money wages, and the same 1,000 workers are employed at the same net wage, $19 in money wages plus the $1 benefit. The workers are no worse off at a wage of $19 with the health insurance than at $20 without the health insurance because the insurance is worth the $1 that it cost in reduced wages. The employer earns no less profit for providing the health benefit.

For a more detailed analysis, consider (following Lee, 1996) a labor market with a typically downward-sloping demand for labor, D, and a typically upward-sloping supply of labor, S, as noted in Figure 11.2. The demand for labor is related to the marginal productivity of workers. The supply of workers reflects the wage in this industry relative to the wage in other industries. Workers will choose to work in this industry as long as the wage they can earn exceeds their opportunities in other jobs. In Figure 11.2, at equilibrium point b, the equilibrium wage is W_1 and the equilibrium quantity of labor demanded and supplied is L_1. Suppose that workers in the market negotiate a health insurance benefit worth $\$z$/hour at that margin, and it costs employers exactly $\$z$/hour to provide. What happens? Employers who were previously willing to pay W_1 per hour for workers will

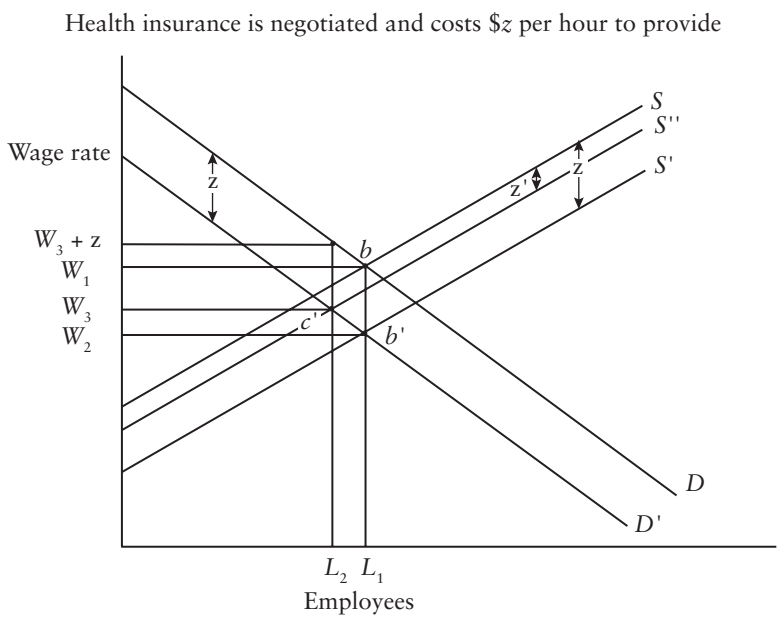


Figure 11.2 Interaction of Health Insurance and Employee Wage

now pay W_1 less $z. Other points on the demand curve will shift downward in a similar manner so the demand curve will shift downward by exactly $z to D. What will happen to the supply curve? Because the workers were willing to supply various amounts of labor at various wage rates according to the supply curve before, now that they are receiving a benefit worth $z they will offer their labor for $z less. Hence, the supply curve will shift downward by exactly $z to S.

What is the result? Similarly to our earlier discussion, the net wage remains the same at W_1, but the money wage falls by $z. The equilibrium wage has fallen to W_2 at point b or by exactly the amount of the benefit. Workers have taken their benefits in lower money wages, and the same number of workers, L_1, is employed at the same net wage. For a real-world example of who pays, see Box 11.1.

There are several reasons that the marginal benefits of the insurance to the employees may fall short of the employers' marginal costs. Some contracts negotiate subsidized coverage for prescription drugs, at a cost to the employer. However, some employees are healthy and do not use prescription drugs. This benefit has no value to them.

In addition, recall from Chapter 8 that for many types of health care, fractional coinsurance lowers the consumers' marginal costs of treatment and leads them to buy more insured care than otherwise. As a result the benefits on average may be worth less to the workers than what they cost the employers to provide. Without moral hazard, prescription drug coverage would simply reduce the cost of drugs to the workers. However, the drug benefit might induce workers to purchase prescription shampoo or prescription cold medicine rather than less expensive over-the-counter brands.

If the average benefit is worth $z′/hour, or less to the workers than the $z/hour that it costs to provide, then the new supply of labor curve, S'', will have fallen by less than the demand for labor (still D, reflecting what it costs to provide the benefit). Equilibrium will be at c',

BOX 11.1

Employers Shift More Health Care Costs to Employees—How You Feel Depends on Where You Sit

The annual *Employer Health Benefits Survey*, released in September 2015, showed that the average general deductible for workers with single coverage totaled $1,077 in 2015, over three times as much as the $303 average in 2006. That deductible has climbed nearly seven times faster than wages over the past five years.

The study also found that 46 percent of workers with single coverage have a deductible of $1,000 or more. That's up from only 10 percent in 2006. Kaiser's study did not measure family coverage deductibles, which can be more complex, but researchers say that those have grown as well.

Kaiser Family Foundation CEO Drew Altman stated, "It's funny, we used to think of $1,000 as a very high deductible, and now it's almost commonplace," he said. "Consumers have much more skin in the game, and that may be fine if you're healthier and don't use a lot of health care. That could be a real problem if you're chronically ill."

How you feel depends on where you sit. In a CNBC report, Kentucky resident Emmett Krall said the annual deductible of $3,500 on his employer-sponsored health insurance made him think about cost more than he wanted to, especially since his 10-year-old son was diagnosed with Type 1 diabetes in 2014. Krall must pay about $200 a month to cover his son's insulin, needles, and pump. "It causes an anxiety and a stress on my part, because I do get stressed about it, and I don't want him to know about it," he said.

In contrast, college professor Bill Cantor saw his premium fall to only $95 a month for family coverage from around $300 since he switched to a high-deductible health plan a few years ago. He uses a health savings account to set aside money for expenses, and he likes how the plan has made him more aware of costs. The 53-year-old said he caught a $200 mistake on a medical bill that he might have missed if insurance had just covered the claim. "I think it would hold down insurance rates more if people thought about their spending," said Cantor.

Sources: CNBC, "Employers Shifting More Health-Care Costs to Employees," www.cnbc.com/2015/09/24/employers-shifting-more-health-care-costs-to-employees.html, accessed February 2, 2016.

Kaiser Family Foundation, "Employer Health Benefits Survey," September 2015, Washington, DC, http://kff.org/health-costs/report/2015-employer-health-benefits-survey/, accessed November 2016.

rather than b, the money wage will be W_3, and the total wage will be $(W_3 + z)$ rather than W_1. Employers will react to the higher gross wages $(W_3 + z)$ by reducing employment, here from L_1 to L_2.

Spousal Coverage: Who Pays?

Working members of the same family often have coverage from several sources. Using the logic from the previous section, an analysis derived from Mark Pauly (1997) helps to

examine the issue of spousal insurance coverage. What happens if the husband has family insurance coverage where he works, and the wife chooses not to take coverage where she works? Who pays in this case? The subtleties of the analysis that occur through the labor market may surprise some readers. Table 11.1 keeps track of the numbers.

Consider a town with 10,000 adults, half men and half women. There are 4,000 married couples, 1,000 single men and 1,000 single women. To simplify, suppose that employees can work in either the Alpha or Beta sector. Alpha employers employ only married men (4,000 employees); half of their spouses (2,000 women) do not work, and half of their spouses (2,000 women) work in the Beta sector. Half of the Beta employees are the spouses (2,000 working women) and half are single (1,000 men and 1,000 women). Thus, each sector has 4,000 workers at the outset; assume that no health insurance is provided and the "pure" wage (W_1 from Figure 11.2) for each employee in both the Alpha and Beta sectors is $80,000 per year. Assume that whatever health insurance is implemented, all employees of each firm receive the same take-home pay regardless of insurance cost.

Suppose now that the Alpha firms (employing only married men) offer to buy family coverage for their employees worth $12,000 per year. The Beta firm (half spouses; half single) offers to pay $6,000 per year per person for their employees, as long as those *who elect coverage* pay an additional $30 per month, or $360 per year (those who do not elect coverage receive no cash in lieu of benefits). In this situation, all Alpha workers will choose family coverage; it is a better buy because the employee is covering himself, his spouse, and his children for $12,000. As a result of market processes similar to those in Figure 11.2, money wages for Alpha workers fall by $12,000 to $68,000. In the Beta sector, wages per worker will fall by $3,000 to $77,000 (because *half* of the workers use coverage that costs $6,000 per person). All told there are 6,000 benefits policies written (because 2,000 married couples share a policy). Who pays?

- *Two-worker* families covered through Alpha firms pay $15,000 for $12,000 in coverage. This occurs because wages have fallen by $12,000 in the Alpha sector (where the men work) and by $3,000 where their spouses work. The 2,000 two-worker couples pay $6,000,000 in total more than they receive.
- *Single-worker* families covered through Alpha firms pay $12,000 for $12,000 in coverage, again because wages have fallen by $12,000 in the Alpha sector. These families pay for exactly what they receive.
- Single-worker households covered through Beta firms pay $3,360 (reduced wages of $3,000 + $360 from the monthly payments) for $6,360 ($6,000 + $360 from the monthly payments) in coverage. The 2,000 recipients pay $6,000,000 in total less than they receive.

The market has transferred income from the working spouses, who do not take the insurance, to the singles, who pay less than it costs for their insurance. The sum of the transfers equals zero, because 2,000 workers (the dually covered spouses) are giving up something that they paid for (in terms of reduced wages), to 2,000 single workers (in Sector Beta) who are being subsidized by the foregone wages of their colleagues.

Are the Beta firms better off because they don't pay for the health care for half of their employees? No, because they are still paying a net wage of $80,000 for the labor services that they use. Are the Alpha firms worse off because they are paying benefits for people who don't work for them? No, they too are still paying a net wage of $80,000 for the labor services that they use. The workers pay for their benefits, but some are subsidized by others!

Table 11.1 Who Works and Who Pays? Alpha and Beta Sectors

a. Who Works?

	Alpha	*Beta*		*Coverage?*
Married Men	4,000	0		Family takes coverage
Married Women	0	2,000		Do not take coverage
Single Men	0	1,000		Take coverage
Single Women	0	1,000		Take coverage
Total	4,000	4,000		
Working in the Home			2,000	Coverage from spouse who works at Alpha

b. Wages?

	Alpha	*Beta*
Initial Wage	$80,000	$80,000
Family Coverage	$12,000	$6,000
Copay per year		$360
Amount on Paycheck	$68,000	$77,000

c. Who Pays?

		Give Up	*Get*	*Number of Policies*	*Total Gain or Loss*
Two Worker	Alpha	$15,000	$12,000	2,000	–$6,000,000
	Beta	None	None		
Single Worker	Alpha	$12,000	$12,000	2,000	$0
Single Worker	Beta	$3,360	$6,360	2,000	$6,000,000
Net transfer					$0

How the Tax System Influences Health Insurance Demand

Since World War II the tax treatment of health insurance has been one of the most important factors in the increased demand for health insurance. Suppose Sara earns $1,000 per week and would like to buy health insurance. Ignoring state and local taxes, assume that she is in the 25 percent marginal tax bracket, so her take-home pay is $750 per week. Suppose further that health insurance would cost her $100 per week. Her net take-home pay would then be the take-home pay of $750 less the health insurance of $100, or $650 per week.

Suppose instead that Sara's employer purchases insurance for her, again at a price of $100 per week. This fringe benefit is exempt from income taxation, as it has been since World War II. Although Sara's total compensation is still $1,000 per week, she is taxed only on the wage portion, or $900. Her take-home pay will now be 75 percent of $900, or $675 per week. Her $25 improvement in well-being occurs because she does not pay $25 in tax on the $100 insurance benefit. The $675 in net compensation with insurance is clearly superior to the $650 net take-home pay without insurance. The tax system has paid $25 per month of her insurance.

If marginal tax rates increase, consumers have incentives to increase employer health expenditures (calculate the numbers above if Sara's income tax rate is 40 percent). Employers also benefit from this arrangement because their levels of Social Security taxes will fall. In 2016 employees pay 6.20 percent for the Social Security portion, and employers pay 6.20 percent. Both groups pay 1.45 percent for the Medicare portion. Because insurance is an expense to the employer rather than a wage (on which Social Security and Medicare taxes must be paid), it is exempt from Social Security and Medicare taxes. In the 1950s, federal marginal income tax rates went as high as 91 percent, and even today many people pay marginal (federal plus state) tax rates of 40 percent or more.

The allocative problem within the economy occurs because health expenditures have been singled out for special treatment. Consider Figure 11.3. This figure shows an entire wage package consisting of the sum of total wages, *W*, and total insurance, *I*. Intercept *M* on the *y*-axis shows the amount of wages if no insurance is in the package. Similarly,

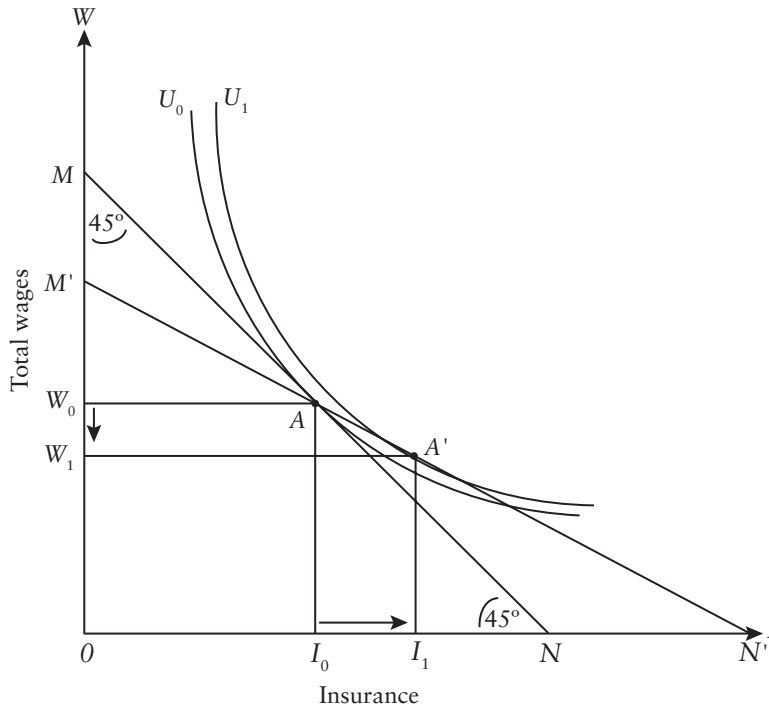


Figure 11.3 Impacts of Preferential Treatment of Employee Insurance

intercept N on the x-axis shows the amount of insurance in the unlikely case that Sara received her entire compensation as insurance benefits. Without special tax treatment, then line MN has a 45-degree relationship to the x- and y-axes, a slope of –1.0. In other words, \$1 of insurance trades for \$1 in wages, and the initial allocation is at point A, with wages W_0 and insurance I_0.

The subsidy of health insurance through the government policies lowers the price of \$1 of insurance relative to \$1 of wage remuneration. Suppose that the employees even recognize the subsidy and are prepared to give up some wages for an insurance subsidy. The "give back" rotates the x-intercept down to M', but the subsidy causes the budget constraint line to rotate to point N' on the x-axis. Hence, \$1 of wages actually buys (trades off for) more than \$1 of insurance. Figure 11.3 shows that without special tax treatment, Sara consumes package A of I_0 and W_0. Even if the "give back" left Sara unchanged at point A, the changed relative prices will now make it more attractive to move to a more insurance-rich package. Thus, the tax system leads Sara to choose combination A', with more insurance at the expense of lower wages, and increased utility at U_1. Not only will Sara buy more insurance, but the tax subsidy may encourage her to insure for the kinds of low- or high-probability events (e.g., routine dental care) that might otherwise be left uninsured.

Who Pays the Compensating Differentials?—Empirical Tests

The compensating differentials, with respect to wages and insurance, merit serious empirical investigation. Many empirical studies have associated health insurance with higher, rather than lower, wages. Because compensation (wages plus insurance) is based on productivity, employers spend considerable effort identifying workers who are better motivated, more dependable, more highly skilled, and better able to interact with clients and customers. Researchers often have had only age or schooling measures to capture productivity and other attributes of more or less productive workers have been unobserved. Because more productive workers get both higher wages and more health insurance, the substitution between wages and insurance may be swamped by the productivity effect. Despite these problems, several researchers have developed creative tests to identify the wage–insurance trade-off.

Gruber and Krueger (1992) examine workers' compensation insurance, and Gruber (1994) looks at mandated maternity benefits coverage. Both studies confirm the existence of "group specific" average wage adjustments. That is, those groups that were paid more in benefits received lower wages.

Jensen and Morrisey (2001) use 1994 and 1998 data on wages, health insurance coverage, demographic characteristics, and health status measures for workers and their spouses from the Health and Retirement Survey (HRS) to examine the wage-coverage trade-off for workers born between 1931 and 1941. They find evidence of compensating differentials for older workers. Other things equal, those workers with health insurance had annual wages about \$6,300 lower than those without. Since annual family health insurance coverage provided through an employer often costs \$6,000 to \$7,000 per year, the evidence suggests that workers do pay for their health insurance through lower wages.

Bhattacharya and Bundorf (2005) look for differentials in the context of obesity. They find that the incremental health care costs associated with obesity are passed on to obese workers with employer-sponsored health insurance in the form of lower cash wages. In their study obese workers in firms *with* employer-sponsored insurance received lower wages, while those *without* employer-sponsored coverage, who had individual coverage, or no coverage, did not.

Adams (2007) examines the impacts of the 1993 New York imposition of pure community rating on firms in the small group market. Community rating prevents carriers from charging different premiums based on age. If smaller firms had been cutting wages of older workers prior to pure community rating to offset their higher health care costs, then one would have expected the reform to lead to higher relative wages for older workers at these firms. The reform did increase the relative wages for older workers, both in relation to older workers in other states and in relation to older workers at large firms within the state.

Emanuel and Fuchs (2008) sum up the trade-off between wages and premiums as "not a point merely of economic theory but of historical fact." Since the late 1970s insurance premiums increased by 300 percent (a factor of four) after adjustment for inflation. Corporate profits per employee flourished, with inflation-adjusted increases of 150 percent before taxes and 200 percent after taxes. In contrast, average hourly earnings of workers in private nonagricultural industries were *stagnant*, actually decreasing by 4 percent after adjustment for inflation. Rather than coming out of corporate profits, the increasing cost of health care resulted in relatively flat real wages for 30 years.

More recently, researchers have examined the trade-off between wages and benefits in the public sector—one in seven U.S. employees work for state and local government. These markets differ from the strictly competitive model in two ways:

1 There is considerable union representation in the public sector, limiting labor supply flexibility.
2 Because the prices for state and local services are related to the tax rate, they are not as flexible as those in the public sector. Tax increases may be directly constrained by law, or may be politically difficult to pass.

Clemens and Cutler (2014) estimate that the compensation of school district employees tended to rise by 85 cents for each dollar increase in health benefits, with reductions in wages and salaries offsetting roughly 15 cents of the increase. They also find that strong public worker organizations can resist offset, possibly at the costs of lay-offs (consistent with the analysis in Figure 11.2). Workers represented by weaker unions face larger offsets.

Qin and Chernew (2014) examine offsets in the larger public sector. Their estimates are consistent with those of Clemens and Cutler (about a 15 percent trade-off). When they exclude health sector employees, and conduct more sophisticated matching analyses, they estimate a trade-off of about 48.5 percent, but it is not statistically different from the earlier 15 percent estimate.

Other Impacts of Employer Provision of Health Insurance

Employer provision of health insurance has other impacts as well. Because the employer is a large, single buyer of coverage, the purchase of insurance through the employer provides scale economies of dealing with insurance providers that single purchasers could never enjoy. This tends to lower the effective price of coverage to the employee.

In addition, group purchase by employers addresses the problem of adverse selection in the provision of insurance. Recall that in Chapter 8 we considered a club whose members participated in an insurance arrangement. The arrangement worked well because the contract provided a necessary service to the members. In particular, the probability of a claim was a random event that could be calculated, and that was independent of the actions of the members.

Central to this result is the proposition that the probability of usage is independent of the insurance plan. Suppose that Karen smokes cigarettes and knows that her probability of a claim is not the 5 percent assumed by the insurance company, but rather 10 percent. If able to convince an insurer that she indeed belonged to the less risky (5 percent) category, Karen would be able to buy insurance much cheaper than the actuarially fair premium. Karen would get a bargain; the insurer would lose money. The inability to identify probabilities, and hence their impacts on the insurance market, is often referred to as *adverse selection*.

As an example, consider an insurance plan that offers major hospitalization coverage. Consider also that many heavy smokers may recognize their higher probabilities of lip, throat, or lung cancer and heart disease. If they can prevent their insurers from finding out about their smoking, then they can purchase much cheaper insurance than the appropriate premium, given their prior conditions.

BOX 11.2

For Many with Pre-Existing Conditions, Obamacare's Flaws are Only a Small Price to Pay

"Fiona O'Connell is familiar with the working person's health care nightmare—the one where you get too sick to work, and then you lose your job, and then you have no insurance to pay for the treatment you need. O'Connell lived that nightmare, and she's still bitter and angry," wrote reporter Judy Peres, in the *Chicago Tribune*, in May 2014.

O'Connell, 54, had worked as a property manager for a company whose benefits included employee medical insurance but in 2007 she received a breast cancer diagnosis and went through months of surgery, radiation, and chemotherapy. After stints on short-term disability and unpaid leave, O'Connell offered to return part time but was turned down. Eventually she lost her job, and her cancer meant she was uninsurable in the private market after her COBRA (Continuation of Benefits, a federal act giving certain former employees, retirees, spouses former spouses, and dependent children the right to temporary continuation of health coverage at group rates—generally for up to 18 months) benefits ran out.

With help from a relative who works in the insurance industry, O'Connell found coverage through the state's former high-risk pool, known as ICHIP. It was expensive—$900 to $1,200 per month, she said—but she felt grateful to have it.

Under Obamacare, O'Connell acquired a Blue Cross Blue Shield policy at a cost of $332.95 per month, including dental coverage (including an income-related federal subsidy of $354.) Her annual deductible is $1,000 for network providers and $2,000 for out-of-network providers. "But most of my doctors were in the plan I chose," she said, "so I have no issues there."

Choosing an insurance plan can be complex. Researchers at Virginia Commonwealth University reported that about 50 percent of consumers buying insurance on mock exchanges picked plans that did not offer adequate coverage for their health status. In addition, consumers should not assume all providers at a given hospital are

covered just because some are. "Each doctor signs their [sic] own contract with an insurer," reported a spokeswoman for Blue Cross and Blue Shield of Illinois. "Even doctors within the same practice may not be in the same network."

To O'Connell and several other cancer patients interviewed, however, any drawbacks are a small price to pay for access to good medical care.

Source: Peres, Judy, "For Many with Pre-Existing Conditions, Obamacare's Flaws Are Only a Small Price to Pay," *Chicago Tribune*, May 29, 2014, http://medcitynews.com/2014/05/pre-existing-conditions-obamacares-drawbacks-small-price-pay/, accessed February 5, 2016.

It can be argued that the purchase of insurance by employers minimizes adverse selection by providing a more appropriate pool for the fixing of insurance rates. These advantages accrue because most groups contain a broad mix of risks, by virtue of having been formed for some purpose *other* than insurance.

Employer-Based Health Insurance and Labor Supply

Because even under Obamacare most private health insurance is obtained through employment and is typically not portable to different employers, researchers have sought to determine the extent to which health insurance may affect labor supply. The two major impacts relate to retirement age and job mobility.

Health Insurance and Retirement

Aging workers face a dilemma. Gruber and Madrian (2002) show that compared with those age 35 to 44, those age 55 to 64 are:

- twice as likely to report themselves in fair health and four times as likely to report themselves in poor health,
- seven times as likely to have had a heart attack and five times as likely to have heart disease, and
- 40 percent more likely to have a prescribed medicine (with twice as many medicines if receiving a prescription).

As a result, their medical spending is almost twice as large and twice as variable as the younger group.

While declining health makes retirement more attractive, it also makes employer-provided insurance more attractive, especially for those younger than 65 years of age, at which time Medicare will provide insurance. Thus, individuals face an incentive to postpone retirement until they are eligible for Medicare at age 65.

Researchers have generally focused on the impact of retiree health insurance on retirement behavior. Gruber and Madrian summarize 16 studies and report that the availability of retiree health insurance raises the odds of retirement by between 30 and 80 percent.

Health Insurance and Mobility

Health insurance may also affect worker mobility between jobs. Prior to the Affordable Care Act, employees may have feared losing coverage for pre-existing conditions, generally defined as any medical problem that has been treated or diagnosed within the past six months to two years. This job lock may have several economic effects:

- Less productive workers may stay at jobs for insurance reasons only, leading to decreased economic output because they would not be replaced by more productive workers.
- Even if all workers are equally productive, some workers may stay in jobs for fear of losing the health insurance benefits to the exclusion of those who would otherwise fill the jobs.
- Those who do change jobs may be denied coverage, face higher premiums, or only obtain insurance subject to a waiver that excludes coverage of their health condition.

Both Cooper and Monheit (1993) and Madrian (1994) address the issue. We look more closely at Madrian's presentation.

Madrian created a simple matrix of the probability of job mobility to consider the impact of job lock. Because job lock is caused by the potential loss of health insurance coverage with changing jobs, one would not expect those with coverage through both their own employment and an outside job to face job lock.

The Probability of Changing Jobs

	Employer-Provided Health Insurance	
	No	*Yes*
No other health insurance	*a*	*b*
Other health insurance	*c*	*d*

She tests for the magnitude of job lock by examining whether those workers with employer-provided health insurance and other coverage are more likely to change jobs than those without alternative coverage, or:

$$(\text{cell } d \textit{ probability}) - (\text{cell } b \textit{ probability}) > 0$$

However, if a man is in cell *d*, it may be due to the insurance provided by his wife, who may be providing income as well; all else being equal, the additional income could lead to increased mobility. Hence, Madrian derives a second test: Whether having other health insurance increases mobility more for those who have employment-based insurance $(d - b)$ than for those who do not $(c - a)$, or:

$$(d - b) > (c - a)$$

This test is referred to as "difference-in-difference," the difference between $(d - b)$ and $(c - a)$.

Her most general model looks at the probability of turnover of married men. Inserting predicted job turnover probabilities into Madrian's matrix, the raw estimate indicates that the mobility rate under job lock (cell *b*) is 0.085 or 26 percent lower than cell *d*, which shows a mobility rate of 0.115. The difference-in-difference estimates, attempting to account for any independent effect of other health insurance on mobility, give an alternative estimate of 31.1 percent.

Subsequent research has supported Madrian's results. Sanz de Galdeano (2006) uses the Survey of Income and Program Participation (SIPP) 1996–2000 panel to find that employer-provided health insurance adversely affects job mobility for all population subgroups by about 31 to 58 percent. She also evaluates the impact of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 on job mobility, and finds evidence that is contrary to its intended objectives. That is, the 1996 HIPAA failed to remedy the insurance-induced job lock in labor markets.

Rashad and Sarpong (2008) provide a good review of the literature. They find that individuals with employer-provided health insurance stay on the job 16 percent longer and are 60 percent less likely to voluntarily leave their jobs than those with insurance that is not provided by their employers.

Job lock thus appears as an unintended consequence of employment-related coverage in the United States. It could be addressed through changes that are broadly consistent with prudent insurance practices. These include elimination of pre-existing condition clauses and the development of health insurance pooling mechanisms in local labor markets that might promote continuity of coverage across employers. The elimination of pre-existing condition clauses due to the features of the ACA would lead observers to expect the degree of job lock to diminish, but as of 2016 there have been no definitive studies.

The Market for Insurance

Having discussed the provision of insurance in theory, we now consider some institutional features of the health insurance market. Such a discussion must address the roles of the Blue Cross insurers, which were originally nonprofit firms, and the commercial insurers, which were typically for-profit firms. With increased competition in the health care sector, many distinctions have blurred. Nonetheless, to understand the current insurance market, we consider how it has developed over time.

The Market for Private Insurance

The number of those privately insured in the United States burgeoned after World War II. From a base of 12 million insured in 1940, it increased by a factor of more than six by 1950 (to 76.6 million), and doubled again (to 158.8 million) by 1970. By 1980, over 187 million U.S. residents had private health insurance.

Well into the 1970s, most of the coverage was provided either by insurance companies (usually in group settings) or by Blue Cross and Blue Shield plans. Since the late 1970s, use of other plans has increased, with declines in both the shares and the absolute numbers of those covered by both the insurance companies and the "Blues." This reflects the movement toward self-insurance by large firms, as well as toward various arrangements through health maintenance organizations (HMOs), preferred provider organizations (PPOs), point-of-service (POS) providers, and other forms of managed care.

Table 11.2 provides more recent health insurance coverage status, as well as type of coverage. The largest portion of the population has private coverage, and the largest portion

Table 11.2 Health Insurance Coverage Status and Type of Coverage, 1990–2012 (in thousands)

Year	*Total People*	*Total Insured Private or Gov't*	*Private*				*Government*			*Not Covered*
			Total	*Employer-Based*	*Direct Purchase*	*Total Insured*	*Medicaid*	*Medicare*	*Military*	
1990	248,886	214,167	182,135	150,215		60,965	24,261	32,260	9,922	34,719
1995	264,314	223,733	185,881	161,453		69,776	31,877	34,655	9,375	40,582
2000	279,517	242,932	205,575	181,862	28,432	68,183	28,062	37,787	8,937	36,586
2001	282,082	244,059	204,142	179,984	28,398	70,330	30,166	37,870	9,580	38,023
2002	285,933	246,157	204,163	179,563	29,287	72,825	31,934	38,359	9,892	39,776
2003	288,280	246,332	201,989	177,362	28,826	76,116	34,326	39,284	10,124	41,949
2004	291,166	249,414	203,014	177,924	29,161	79,480	38,055	39,757	10,584	41,752
2005	293,834	250,799	203,205	178,391	28,980	80,283	38,191	40,167	11,164	43,035
2006	296,824	251,610	203,942	178,880	29,033	80,343	38,370	40,336	10,543	45,214
2007	299,106	255,018	203,903	178,971	28,500	83,147	39,685	41,387	10,955	44,088
2008	301,483	256,702	202,626	177,543	28,513	87,586	42,831	43,031	11,562	44,780
2009	304,280	255,295	196,245	170,762	29,098	93,245	47,847	43,434	12,414	48,985
2010	306,553	256,603	196,147	169,372	30,347	95,525	48,533	44,906	12,927	49,951
2011	308,827	260,214	197,323	170,102	30,244	99,497	50,835	46,922	13,712	48,613
2012	311,116	263,165	198,812	170,877	30,622	101,493	50,903	48,884	13,702	47,951

Sources: 1990 and 1995 data from *Current Population Survey*, March 1988–2005; 1999 data forward from www.census.gov/hhes/www/hlthins/data/historical/files/hihistt1.xls, accessed November 2016.

of that comes through the workplace. In 2012, 198.8 million people had private coverage, with almost 86 percent (170.9 million) getting the coverage from the workplace. Over 101 million had health insurance through the government—some (particularly those with Medicare) were dually covered by both the private market and the government. Based on these government estimates, the number of uninsured increased from 36.6 million in 2000 to 48.0 million in 2012, and touching 50.0 million in 2010.

The period from 2007 to 2009 represents a major departure from longer-term trends. During the "Great Recession," the number of Americans with employer-provided insurance fell by 7.8 million, and the number with government-provided insurance rose by over 10 million, with an over 8.2 million increase in those receiving Medicaid. The number of uninsured jumped from 45.7 million to 50.7 million. Almost certainly, the passage of the 2010 Affordable Care Act stemmed in large part from perceived problems in insurance coverage, as well as concerns about health costs. We discuss this in considerable detail in Chapter 22.

Insurance Practices

At least two organizational and practice issues characterize the health insurance industry and link it to the emerging issues of managed health care. The first issue is a conflict between insurers and the insured (most often represented by the health care providers) regarding the amounts of claims, and indeed whether the claims should be paid at all. From the earliest instances of health insurance, providers, most particularly physicians, argued that their judgments must not be questioned on cost grounds. Insurers, in contrast, could increase their profits and reduce customer premiums by judiciously questioning treatments and costs.

Why would providers consent to having someone second-guess their decisions? Goldberg and Greenberg (1977) traced the growth of health insurance in Oregon in the 1930s. At that time, physicians shared in economic problems of the Great Depression with the larger population, and they saw acceptance of health insurance, even with its accompanying oversight, as a way of increasing earnings. Although insurance plans were attractive to physicians in those difficult times, groups such as the Oregon State Medical Society threatened to expel physicians who participated in the plans. They sought to establish their own plans, and the plans that they established tended to be less strict in their cost reviews.

Blue Cross and Blue Shield plans were started by medical providers, with Blue Cross providing hospital payment and Blue Shield providing physician payment. The Blues typically offered more complete and more comprehensive coverage than other insurers, and they paid participating providers directly. They were also generous in the payment of hospital care. The development and success of large hospitals in the post-World War II United States was at least in part supported by the generosity of Blue Cross and Blue Shield reimbursement of hospital stays on per diem bases. The fiscal distress facing many large hospitals since the early 1980s may be traced to changes in financing that accompanied the reduced power of Blue Cross and Blue Shield leadership in the provision of health insurance, in part related to the strictures placed by Diagnosis Related Group (DRG) controls. Medicare introduced DRGs to provide prospective fixed payments for specific diagnoses. These payments induced cost-containment measures that generally resulted in reduced hospital stays.

The second organizational pattern that has characterized the industry is the change in "rating" clients. We have noted that insurers pool their clientele to determine risk premiums based on their experiences with the groups. Blue Cross plans began with a method called *community rating* in which all subscribers in a given location, irrespective of age or health experience, were charged the same premium. This contrasts with experience, or risk, rating,

in which the insurer charges group premiums (to a company or a fraternal or service organization) based on its experience with the group.

Community rating provides a fundamental information problem. Low-risk clients are overcharged, and their premiums are transferred to higher-risk clients in the same pool. Advocates of community rating argue that this allows high-risk and low-income clients to buy insurance that would otherwise have been unavailable. This argument, in part, characterizes the *individual mandate* from the Affordable Care Act which requires individuals either to purchase health insurance or to pay a penalty (functionally equivalent to a tax) instead.

Under risk pooling, however, insurers can identify groups with low risk and offer them lower-cost insurance. Many large firms choose alternative carriers who, again, will charge lower premiums than the community-rated Blue plans. The large firms may in fact choose to self-insure. This "cream skimming" practice may leave the plans that continue to community-rate their clientele with client pools that are so risky that they require almost prohibitively high premiums.

Health Insurance Markets Since the 1980s

The shift toward managed health care through HMOs and PPOs from about 1980 onward induced a change in philosophy among Blue Cross and Blue Shield insurance plans. The typical Blue Cross and Blue Shield system faced stiff competition from smaller organizations that were peeling off supposedly more desirable clients from the community-rated system leading to higher costs for those who remained in the system.

The major explicit response to the changing marketplace was the 1994 approval of a change in organizational status. Blue Cross plans are now either for-profit firms or establish for-profit subsidiaries.

Why change? One fundamental reason involved the ability to raise capital. Nonprofit organizations must generate funds through their revenues, whereas for-profit firms may sell stock. With more competition in the marketplace and renewed emphases on lowering premiums and costs of care, the option to sell stock became more attractive.

A 2016 report (Farrah, 2016) indicated that Blue Cross and Blue Shield companies and affiliates commanded more than 60 percent of the health insurance market share in ten states, up from nine states at the same point the previous year.

In ten other states, Blues companies collectively retained a 50 to 59 percent share of the market and in another 14 states the market share ranged from 40 to 49 percent. For 17 states plus the District of Columbia, Blue Cross and Blue Shield business represents market shares ranging from 20 percent to 39 percent.

Austin and Hungerford (2010) argue that Blue Cross plans were originally designed to avoid competition by requiring exclusive territories and barring plans linked to specific hospitals. They surmise that those requirements may have been aimed at supporting community rating policies and broadly based risk pools, benefitting many consumers. As commercial insurers and managed care strategies rose in prominence, market forces along with merger and acquisition strategies have helped reshape the health insurance market.

The Uninsured: An Analytical Framework

With the high costs of health care and the inevitability to most people of incurring at least some expenses, it has become crucial for individuals to have access to health insurance.

In the United States, health insurance availability has been linked to the workplace. Yet various surveys showed that leading up to and shortly following the Great Recession of 2007–2009 over 50 million Americans have no health insurance at any moment in time (see Box 11.3), and that a large fraction of these people were employed.

The workplace did not provide insurance for all families, and coverages vary widely by economic circumstance. In 2009, 28.8 percent of those aged 25–34 were uninsured, compared to 23.8 percent in 2002. In the 35-to-44 range, 21.8 percent were uninsured in 2009, compared to 17.8 percent in 2002.

BOX 11.3

Counting the Uninsured

Estimates of the uninsured come from *surveys of the population*, rather than by censuses, in which all are counted. Most estimates of the uninsured come from the National Health Interview Survey (NHIS), the March supplement to the Current Population Survey used by the Census Bureau (CPS), the Medical Expenditure Panel Survey (MEPS), and the Survey of Income and Program Participation (SIPP). Because the surveys differ, their results also differ.

The CPS identifies individuals as uninsured if they have lacked coverage for the entire previous calendar year (although many analysts believe that respondents incorrectly provide information about their current insurance status). The SIPP identifies individuals who are uninsured for each month of a four-month reference period. The NHIS identifies individuals as uninsured if they lacked coverage in the month prior to the survey. The MEPS data count as uninsured those without coverage for the entire interview round (an average of three to five months).

Insurance definitions vary. MEPS defines private insurance as coverage for hospital and physician services, thereby eliminating serious and dread disease, workers compensation, accident, and disability policies from counting as coverage. As well, MEPS has not included single service plans (such as dental plans) as private insurance except for single service hospital coverage. CPS and SIPP instruct interviewers not to count single service plans as private insurance, but researchers believe that some single service coverage may get misreported as comprehensive coverage.

In short, estimating the uninsured is like measuring temperature or rainfall; estimates may vary depending on when, where, and how the measurements are taken. That said, the uninsured estimates from different sources generally move up and down together, serve as checks on the others, and provide reliable trends, but they are not likely to be identical.

Source: U.S. Department of Health and Human Services, "Understanding Estimates of the Uninsured: Putting the Differences in Context," https://aspe.hhs.gov/basic-report/understanding-estimates-uninsured-putting-differences-context, accessed February 6, 2016.

Repeated surveys of the uninsured (Kaiser Foundation, 2005) exploded common myths about the numbers and motivations of the uninsured. Although it is plausible that some people rationally "choose" not to have insurance, only 7 percent asserted that they did not think they needed it. The majority said the main reason was that it was too expensive. Some

members of the public, and some politicians, argue that the uninsured can get the care they need, including emergency room settings, when they really need it. However, surveys indicated that over one-third of the uninsured reported needing care in the previous year but not getting it, and nearly half of the uninsured reported postponing care—rates at least three times higher than those with insurance. Box 11.4 updates the discussion to the most recent 2010 health care debate.

Under any circumstance, it is important to examine the reasons that some individuals lack insurance. Insurers must be able to lower the loading factors, which are the costs of determining probabilities of claims and processing claims. The ability to insure through the workplace gives the opportunity to improve the experience rating. In principle, private insurers can insure those outside the workplace, and many do, particularly those in affinity groups, such as organizations and clubs, or the elderly. If high costs lead to onerous payments, then the problem may be poverty rather than high prices. This would suggest the need for governmental subsidies in a social insurance scheme.

BOX 11.4

Why Being Insured Matters

In the debate about the Affordable Care Act of 2010, there was a heated discussion about the number of uninsured, whether they were uninsured by choice, and what the impact was on their health. A 2010 report by the Centers for Disease Control addressed these issues:

1 More than one in four adults 18–64 years old—about 50 million people—had no health insurance for at least part of the previous year. Over the previous several years, the number of adults 18–64 years old without health insurance for at least part of the year had increased by an average of 1.1 million people each year.
2 Middle-income people accounted for half of that increase.
3 Not having insurance has a greater impact among those adults who need health care the most. Delays in receiving health care can lead to poorer health and higher medical costs over time, especially for those individuals who already have health issues, including the approximately 40 percent of the U.S. population with one or more chronic diseases. About 60 percent of adults ages 18–64 with a disability who had recent gaps in the past 12 months in their insurance skipped or delayed care as well.

Source: Centers for Disease Control and Prevention, "Health Care: See Why Being Insured Matters," www.cdc.gov/Features/VitalSigns/HealthcareAccess/, November 2010, accessed May 24, 2011.

The Working Uninsured

Although the health insurance environment has changed considerably with the passage of the Affordable Care Act, employer-provided insurance is still the biggest source of health insurance. However, some employers cannot or will not offer insurance. What are the economics behind this decision?

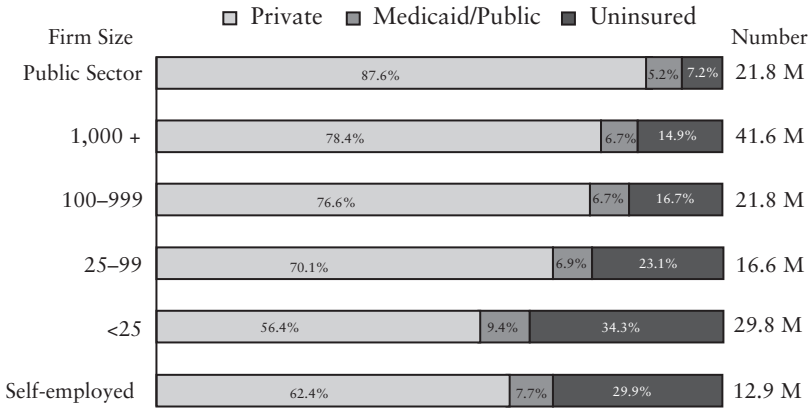


Figure 11.4 Health Insurance Status of Workers by Firm Size, 2009

Source: Kaiser Commission on Medicaid and the Uninsured/Urban Institute analysis of 2010 ASEC Supplement to the CPS.

Consider the working uninsured. Starting in Figure 11.5, the firm that is a price taker (facing a horizontal supply curve) in the labor market pays wage w_0, and offers no insurance. At initial equilibrium point A, the firm hires L_0 workers. Suppose that the workers negotiate an insurance contract worth \$1 per hour to them. As in Figure 11.2 earlier in this chapter, if the workers were willing to accept wage w_0 before, they will now be willing to accept $w_0 - 1$ dollars per hour in wages plus the insurance, and their supply curve shifts downward to reflect this willingness.

The literature suggests, however, that many businesses, especially small businesses, may have limited experience ratings; as a result, their loading costs may be 40 percent or more higher than those paid by large firms. If the loading costs are high, then it may cost much more than one dollar to provide a dollar's worth of insurance. In the Figure 11.5 example, the demand for labor curve (with \$1 per hour of insurance) shifts down by \$1.40.

At employment level L_0, the marginal cost to the firm of the workers, $w_0 - 1$, exceeds the marginal product net of the health insurance, $w_0 - 1.40$. To continue employing L_0 workers, the firm must reduce the amount of insurance to less than \$1 per hour so that the sum of the value of marginal product plus the insurance equals w_0. If it cannot reduce the amount of insurance, then the firm will have to reduce its employment to L_1, at point B.

The Impacts of Mandated Coverage

As of 2016 the 50 states plus the District of Columbia have almost 1,100 mandates to provide specified benefits. The states vary their detail from Idaho (four mandated benefits) to Maryland (63 mandated benefits). Many of the mandated benefits are quite common (mammogram and prostate screening, alcohol and smoking cessation), but others are less so (wigs in Rhode Island, bone mass measurement in Maryland).[1]

The ACA has mandated a set of ten categories of essential benefits at the federal level. These benefits have moved much of the "action" on mandated coverage from the state to the federal level, and the mandated federal benefits have come to dwarf state mandates in their comprehensive nature, and their economic impacts.

The mandated essential benefits are:

1 Outpatient care—the kind patients get without being admitted to a hospital.
2 Trips to the emergency room.
3 Treatment in the hospital for inpatient care.
4 Maternal care before and after the baby is born.
5 Mental health and substance use disorder services: This includes behavioral health treatment, counseling, and psychotherapy.
6 Prescription drugs.
7 Services and devices to help patients recover if injured, have a disability or chronic condition. This includes, but is not limited to, physical and occupational therapy, speech-language pathology, and psychiatric rehabilitation.
8 Laboratory tests.
9 Preventive services including counseling, screenings, and vaccines to keep patients healthy and care for managing chronic diseases.
10 Pediatric services including dental care and vision care for children.

Source: "10 Health Care Benefits Covered in the Health Insurance Marketplace," www.healthcare.gov/blog/10-health-care-benefits-covered-in-the-health-insurance-marketplace/, accessed February 7, 2016.

Returning to Figure 11.5, if such coverage is provided through employee insurance, the mandates would raise the amount of insurance as well as its costs, thus lowering the (dashed) demand for labor curve net of insurance. Responses of employers to the increased marginal costs brought on by mandates may result in two adverse impacts. First, the company may stop offering insurance entirely because it is too expensive. Thus, rather than having modest health coverage with the benefits of whatever experience rating may exist within the workplace, there may be no coverage at all. The employer then may have to raise the wage to keep employees who would have to buy their own insurance. This may allow for employee choice, but it also denies the employee workplace-related experience rating.

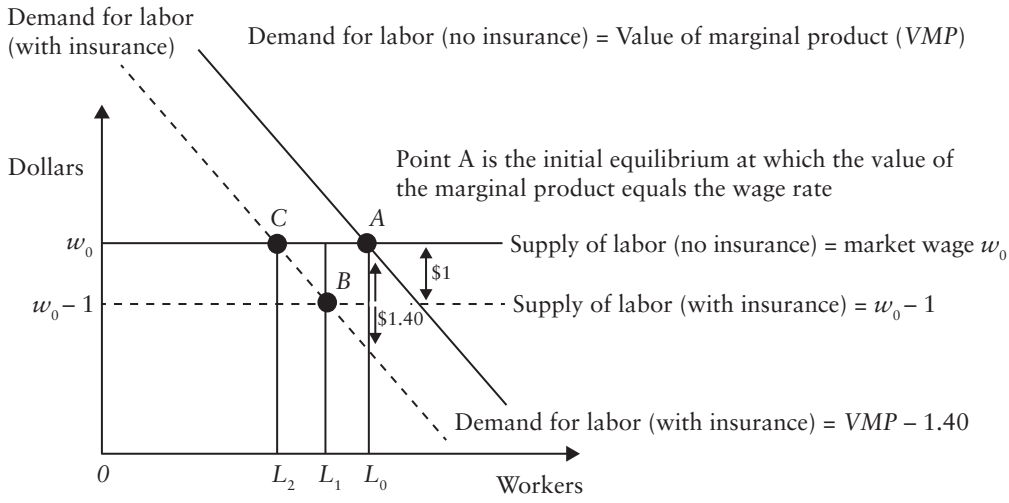


Figure 11.5 Insurance and Employment

The second adverse impact may also be understood by examining Figure 11.5. The equilibrium value of marginal product and net wage at point *A* represents a labor force of the appropriate size to maximize profits for the producer. Suppose, for example, that w_0 represents a binding minimum wage, either by federal or by local "living wage" laws. Even if the workers value $1 in mandated benefits by the full dollar and would be willing to reduce their supply price, the money wage *cannot* fall lower than w_0. By this analysis, the new equilibrium is at point *C*, and the mandated coverage is likely to result in additional unemployment by the amount $(L_0 - L_2)$, which of course means sharply curtailed insurance benefits. How much employment falls is related to the elasticity (or responsiveness) of the labor demand curve to the increased gross wage.

Many ACA opponents bolster their position with this "job killing" argument. However, the benefits of the enhanced coverage must be weighted against potential costs. In addition, although it has potential theoretical merit for an individual firm (where no other firm is subject to the mandate), it is unclear how the mandate will impact the entire economy, where all firms are faced with mandates.

Impacts of the Affordable Care Act on the Uninsured

The ACA's major coverage provisions went into effect in January 2014 and led to significant coverage gains. The number of uninsured nonelderly Americans in 2014 was 32 million, a decrease of nearly 9 million since 2013. We will discuss the details of the ACA in considerable depth in Chapter 22, but since one of the major goals of the ACA was to reduce the number without insurance, it seems most appropriate to map out the enrollment successes (as of this writing in 2016), which include substantive reductions in the numbers uninsured, as well as the characteristics and issues that have as yet left large numbers of Americans without health insurance. This analysis draws on U.S. Census data analyzed by the Kaiser Commission on Health and the Uninsured (2015).

Elements of the ACA

The ACA has had two major means by which the uninsured can get coverage. The first, and initially the more publicized, was the establishment of health exchanges (think of an Orbitz® for health insurance). On these exchanges, health insurance carriers would compete for customers. The improved information, and supposed ease in enrollment, would enable health insurance consumers to shop among plans (termed platinum, gold, silver, and bronze, based on their coverage and cost).

The second important element was the subsidized expansion of Medicaid funding for those households whose incomes were below, to up to 400 percent of the federal poverty level (FPL). Although the details are nontrivial, the important fact is that this improved purchasing power would allow households to buy insurance for the first time and/or reduce the amount that they were paying.

Evidence on the Impact of the ACA on the Uninsured

While undoubtedly some of the fall in the number of uninsured reflected a continued recovery from the most severe recession since the 1930s, both the numbers and percentages

of those without health insurance have fallen to the lowest levels in at least 25 years, and the percentages of those without health insurance have fallen to the lowest since records have been kept.

As of June 2015, Medicaid enrollment had grown by 14 million since the period before open enrollment (which started in October 2013). This growth constituted an increase of 23 percent in monthly Medicaid enrollment. This enrollment increase corresponded with large declines in the uninsured rate. As noted in Figure 11.6, between 2013 and 2014, the uninsured rate dropped significantly, from 16.2 percent in the last quarter of 2013 to 12.1 percent in the last quarter of 2014. Declines continued into 2015, with preliminary data indicating an uninsured rate of 10.7 percent. Children, who already had a low uninsured rate due to relatively higher eligibility levels for public coverage, experienced a small decline in the uninsured, while the uninsured rate among nonelderly adults dropped significantly. The data also show substantial coverage gains among poor and low-income individuals, and people of color.

Who was still uninsured? Figure 11.7 shows that of the 32.3 million still uninsured, 73 percent of the uninsured had one or more full-time workers. Individuals below the poverty level ($19,055 for a family of three in 2014) were at highest risk of being uninsured. Over 80 percent of the uninsured were in low- to moderate-income families. People of color were at the highest risk of being uninsured, although the largest numbers were non-Hispanic whites.

What was the major cause for those who were uninsured? Costs were the biggest reason. In 2014, almost 50 percent of uninsured adults said that they lacked coverage because it was too expensive. Another 25 percent said they were ineligible because they were unemployed, could not get offers through work, or were told that their immigration status rendered them ineligible.

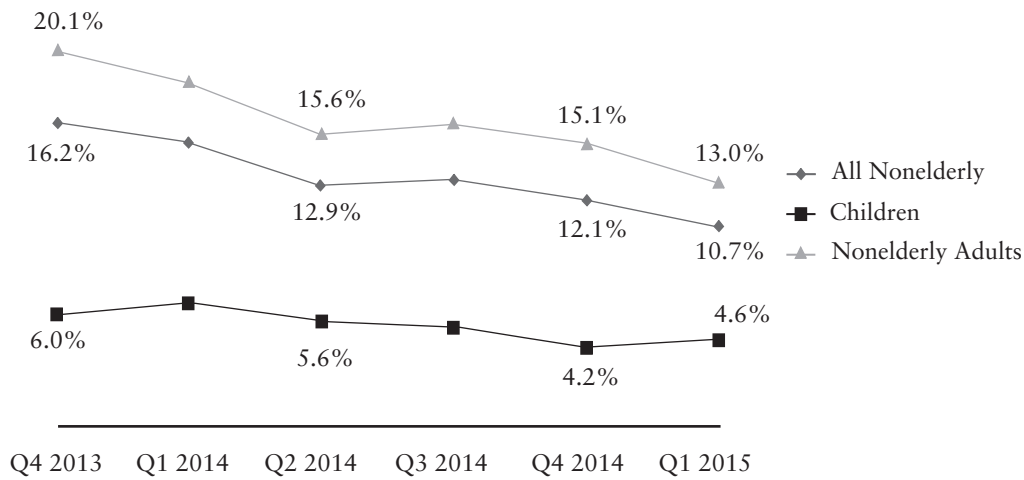


Figure 11.6 Quarterly Uninsured Rate for the Nonelderly Population

Source: Urban Institute estimates based on 2008–2010 Medical Expenditure Panel Survey.

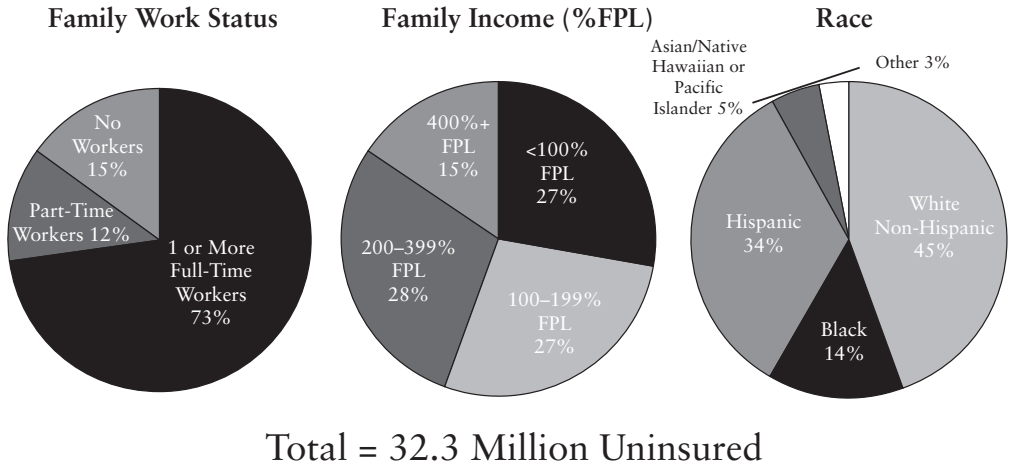


Figure 11.7 Characteristics of the Nonelderly Uninsured, 2014

Note: The U.S. Census Bureau's poverty threshold for a family with two adults and one child was $19,055 in 2014. Data may not total 100% due to rounding.

Source: Kaiser Family Foundation analysis of the 2015 ASEC Supplement to the CPS.

Conclusions

Chapter 8 introduced readers to insurance and to the specific issue of health insurance for individuals. In the current chapter, we have focused on the insurance market and the behaviors of firms within that market. We established that in a market setting, insurance constitutes an important part of the wage package, and to the extent that it is valuable to the workforce, higher insurance is reflected in lower money wages. This market result occurs irrespective of who contractually pays for the insurance.

We have also shown how many of the trappings of the U.S. health care system are related to the employer base of the health insurance. The system of tax deductibility pushes employees to ask for higher proportions of untaxed insurance relative to taxed wage benefits. The linkage of insurance to the workplace also tends to lock employees into certain jobs, constraining mobility.

We continued with a brief discussion of the evolution of the health insurance industry. This industry was formerly defined by the Blues (Blue Cross and Blue Shield) with their nonprofit status, community rating, and predilection for hospital care. Although they are still large, the decline in the primacy of the Blues has led to profound changes in the provision of health insurance and the delivery of care.

The chapter also included several implications about the uninsured. Some are not employed and hence ineligible for health insurance. There are others, however, whose health, employment, or lifestyles may not permit commercial insurers to provide insurance profitably. Government mandates to employers that they insure everyone if they insure anyone may cause employers to drop insurance plans entirely, thus leaving larger numbers at risk.

Finally, the ACA, while hardly a complete solution to problems of the uninsured, has changed the landscape in which large-scale and universal coverage can be viewed.

Summary

1 Individual health insurance, in theory, trades off a guaranteed reduction in wealth (the insurance premium) for a reduction in uncertainty due to ill health through the pooling of risk. The organization and cost conditions in some health insurance markets, however, suggest that some contingencies may not be insurable.
2 The economies of scale in processing information suggest that smaller firms or unions may see high marginal insurance costs relative to marginal benefits levels. These higher costs may reduce or eliminate the range of services that they offer.
3 If workers in an industry value health insurance, then competitive pressures lead to reduced money wages' offsetting increased health benefits.
4 Subsidizing health insurance through government tax policies lowers the price of $1 of insurance premium relative to $1 of wage remuneration. This leads employees to purchase more health insurance relative to wages than would otherwise occur.
5 Health insurance is a key determinant in the decision to retire. Studies suggest that the availability of retiree health insurance raises the odds of retirement by between 30 and 80 percent.
6 Employer-provided health insurance may inhibit worker mobility between firms, thus locking employees into jobs. Researchers find this job lock to be responsible for substantial reductions in employee mobility.
7 Much of the American health insurance environment has been defined by Blue Cross and Blue Shield plans, which as nonprofit firms were typically exempted from profits taxes, property taxes, and federal and state corporate taxes. Analysts once felt that such tax treatments gave the Blues considerable competitive advantages, but over the past two decades the Blues have faced considerable competition.
8 Many groups have advocated that the federal government or individual states mandate either more coverage or various types of coverage. In reaction:
   - Companies may stop offering insurance entirely because it is too expensive.
   - Marginal workers may now cost more than they are worth, and some of them will be let go.
9 The ACA introduced a set of 10 essential benefit categories, and mechanisms for increasing insurance coverage. Although open enrollment began only in October 2013, preliminary indicators show dramatic decreases in the numbers of uninsured Americans.

Discussion Questions

1 Suppose each person's health expenditures can be predicted with certainty by both the insured and the insurer. What are the implications for insurance markets? Explain the prevalence of insurance for highly predictable events, such as routine dental services.
2 In 1986, the U.S. federal income tax system changed marginal tax rates so that the top federal marginal rate fell from 50 to 33 percent. From what you know about how fringe benefits are negotiated, what would you expect to happen to the demand for health insurance as a fringe benefit? Why?

3 Suppose your health insurance allows you, a worker, to buy whatever prescription drugs you wish for $5 per prescription. In contract negotiations it is proposed to change this benefit to "10–20," that is, you pay $10 for generic drugs and $20 for brand-name drugs. What would your reaction be? What would economic analysis predict?
4 Suppose that a company pays its workers $20 per hour and provides an additional $2 per hour worth of fringe benefits, including a basic health insurance policy. Discuss the firm's reaction to a state mandate that requires it to expand the items covered in the health care policy. What is likely to happen to the number of people employed?
5 Blue Cross plans typically have practiced community rating. If other insurance firms are seeking healthier patients at reduced rates, what impact will this have on Blue Cross net revenues? Why?
6 According to the ACA, adults with pre-existing conditions became eligible to join a temporary high-risk pool, which will be superseded by the health care exchange in 2014. To qualify for coverage, applicants must have a pre-existing health condition and have been uninsured for at least the past six months. Analyze the impacts of such regulations on insurance markets.
7 What is job lock? Would you expect job lock to increase or decrease if employer-based health insurance were to be replaced by government-provided health insurance?
8 Suppose a household does not carry health insurance. Can we conclude that this reflects failure of insurance markets? Why or why not?

Exercises

1 Using Figure 11.1, illustrate the probability that someone will obtain insurance for treatment for
   (a) A hangnail.
   (b) A broken arm.
   (c) A "bad hair" day.
   (d) Viral meningitis.
2 Using Figure 11.2:
   (a) Calculate an initial labor market equilibrium (wages and employment) determined by the demand and supply of labor.
   (b) Indicate the wage and employment impacts of a health insurance policy that costs $2 per hour to employers and is worth $1.50 per hour to the workers.
3 Using Figure 11.2:
   (a) Calculate an initial labor market equilibrium (wages and employment) determined by the demand and supply of labor.
   (b) Indicate the wage and employment impacts of a health insurance policy that costs $1.50 per hour to employers and is worth $2 per hour to the workers.
4 Using Figure 11.2, consider an insurance policy that provides free "purple aspirin" to all workers. This benefit provides no conceivable advantage (workers don't care whether their aspirin is purple or white) but comes with cost z. Show the new labor market equilibrium indicating the wage and employment impacts.
5 Using Figure 11.2, indicate the wage and employment impacts of a health insurance policy that costs $2 per hour to the workers and is worth exactly $3 per hour to the workers. Why do your answers to exercises 2 through 5 differ?
6 Consider a difference-in-difference model of job lock. The research finds the following probabilities:

The Probability of Changing Jobs

	Employer-Provided Health Insurance	
	No	*Yes*
No other health insurance	0.15	0.10
Other health insurance	0.25	0.24

(a) Interpret each element of this matrix in terms of the probability of changing jobs.
(b) Does the presence of alternative insurance mitigate job lock? Explain your answer.
(c) Does the difference-in-difference calculation validate the larger impact of "other insurance" found by Madrian? Why or why not?

7 Consider the market labor demand L_D and labor supply L_S, where W is the market wage.

Demand: $L_D = 1{,}000 - 20W$
Supply: $L_S = -200 + 400W$

(a) What is the equilibrium market wage? What is the equilibrium employment level?
(b) Calculate the equilibrium market wage and employment level if the workers negotiate a benefit worth \$1 that costs the employers \$2.
(c) Calculate the equilibrium market wage and employment level if the workers negotiate a benefit worth \$2 that costs the employers \$1.

8 Consider two workers, Ralph and Steve. Both of them work for the same employer, and each earns \$15 per hour. Steve is taxed at the 15 percent marginal rate. However, Ralph is married, and due to his wife's income, he is taxed at the 28 percent marginal rate. Using Figure 11.3, indicate which one would be expected to seek more health insurance and why.

9 Suppose that Charlie's Pizzeria in Kalamazoo, Michigan, employs 10 employees at a wage level of \$9 per person. All other costs (ovens, rent, advertising, return to capital) total \$50 per hour, and the pizzeria sells 15 pizzas per hour at a cost of \$10 per pizza. Suppose there is mandated coverage that can only be covered at a cost of \$1.50 per hour, if it is offered at all. Charlie finds that if he offers insurance, he could maintain production by letting one worker go and running his pizza ovens a little hotter, leading to costs of \$55 per hour.
(a) What are Charlie's original profits?
(b) What is Charlie's elasticity of demand for labor? How is this calculated?
(c) What will happen to Charlie's profits in the short run if he chooses to pay for mandated insurance?
(d) What will Charlie's long-run decision be? Why?

Note

1 The Center for Consumer Information & Insurance Oversight, "Information on Essential Health Benefits (EHB) Benchmark Plans," www.cms.gov/cciio/resources/data-resources/ehb.html, accessed November 2016.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 12

Managed Care

In this chapter

- What Is the Organizational Structure?
- What Are the Economic Characteristics?
- The Emergence of Managed Care Plans
- Development and Growth of Managed Care—Why Did It Take So Long?
- Modeling Managed Care
- How Does Managed Care Differ?—Empirical Results
- Growth in Spending
- Competitive Effects
- The Managed Care Backlash
- Accountable Care Organizations (ACOs)
- Managed Care and the Affordable Care Act
- Conclusions

The previous chapter described how conventional health insurance will generally increase consumers' health care utilization. Those who are insured consider the out-of-pocket cost of care rather than the true full cost at the point of service. In the absence of restrictions, they will purchase services beyond the point at which the marginal benefit of the care equals its marginal cost. For insurance to provide a net benefit to society, the costs of this increased health care consumption must be overcome by the benefits of the reduced financial risk to patients. Even with the benefits of risk reduction, improved insurance coverage leads to increased costs to society.

A simple analogy may help address the dilemma. Suppose that rather than health care insurance, employers provided food and clothing (F&C) insurance for their workers. A "fee-for-service" F&C plan would allow the consumers to purchase their food and clothing at any merchant they choose and would reimburse the consumers subject to coinsurance rates and deductibles. A consumer facing a 20 percent coinsurance rate could purchase filet mignon for $20 per pound and have the insurer pay $16 per pound, or 80 percent of the price. Another consumer could purchase designer athletic shoes for $200 and have the insurer pay $160, again 80 percent of the price. Consumers would likely buy more (or more expensive) filet mignon or athletic shoes than if they had to pay the full amount themselves. The market effect of such plans would likely cause consumers and their insurers to worry about F&C cost and expenditure inflation.

Suppose that in response to the perceived high costs of food and clothing, and the consequent high cost of the F&C insurance plan, a group of consumers and their employers organized and offered a "managed F&C" plan. In this plan, members (consumers and the employers) would pay a fixed amount per person per month for food and clothing, presumably less than they were paying (together) under the fee-for-service plan. In return for this reduced cost plan, the consumers would be limited to shopping at a single shopping center with which plan managers had negotiated lower prices for food and clothing. Moreover, the plan managers could limit the types of goods purchased (no filet mignon) and might also attempt to curb total consumer expenditures in other ways. It is likely that at least some consumers would find such a plan attractive due to its lower costs.

Analysis of this "managed care" F&C arrangement raises a host of questions. For individual consumers, one might ask:

- Are they getting the same quality of goods as before?
- Are they being denied goods that they "should" be getting?
- Are their expenditures reduced?
- Are they less well-fed, less healthy, or less well-dressed than before?

At the market level, one might ask:

- Do aggregate F&C expenditures decrease or does their growth rate decrease?
- Do the managed F&C plans meet consumer preferences?
- Can the merchants earn sufficient returns to stay in business?
- Is there competition in the managed care F&C market, and do the managed care F&C plans influence the fee-for-service F&C plans in terms of the prices or selection of goods?

With this in mind, we turn to the issue of managed health care. One might argue that physician practice must be managed in order to address high health care costs. This suggests that networks of providers, including HMOs (health maintenance organizations), PPOs (preferred provider organizations), and individual practice associations (IPAs), are widely seen as means to restore competition to the health care sector and as means to control

expanding heath care costs. We devote this chapter to managed care with particular attention to the distinctive combination of insurance and care exemplified by HMOs and similar organizations. Unless distinguishing the individual types of institutions, we will refer to them as managed care organizations (MCOs). The HMO receives special attention in this chapter for its pioneering role and for the fact that much of the scholarly and policy research has focused on HMOs. Accountable Care Organizations (ACOs), the Affordable Care Act (ACA), and other recent developments impacting managed care are taken up later in this chapter.

HMOs appear to overcome the information problems inherent in fee-for-service (FFS) health care markets that ordinary insurance coverage may exacerbate. Under FFS, the provider provides health care and advises the consumer on how much to get. At first glance, it appears that the consumer's imperfect information about health care, when combined with FFS remuneration, may create the incentives for substantial overconsumption. The HMO organizational form appears to eliminate the overconsumption incentives and replace them with cost-control incentives and even possibly incentives toward underconsumption.

We begin this chapter by describing managed care and its cost-cutting potential. We then turn to HMOs as a form of health care organization that combines the functions of insurance and the provision of care. We describe the HMO and its organizational relatives, and we assess the theory and evidence on their effects. We then turn to the market effects of managed care on providers, insurers, and the adoption of new health care technology.

What Is the Organizational Structure?

It is instructive to provide a general description of MCOs, leading to a more specific discussion of HMOs, while recognizing that the concept of managed care is undergoing constant changes. Analysts speak of an organized delivery system as a network of organizations (e.g., hospitals, physicians, clinics, and hospices) that provides or arranges to provide a coordinated continuum (from well care to emergency surgery) of services to a defined population. This system is held clinically and fiscally accountable for the outcomes and the health status of the population served. It is tied together by its clinical (treatment) and fiscal (financial) accountability for the defined population. Often the organized delivery system is defined by its association with an insurance product.

In principle, managed care creates incentives for keeping people well by emphasizing prevention and health promotion practices, and by treating those who become ill at the most cost-effective location in the continuum of care. Through a more centralized management of services, the goal is to provide additional quality-enhancing features for a given price, or to provide a given set of quality attributes or outcomes for a lower price. A primary provider typically serves as the patient's gatekeeper to help ensure appropriate care and limit overutilization.

Two features characterize the contemporary MCO. The first is the extensive reliance on health care information systems. Initially, these systems were developed mainly to replace clerical functions such as billing and record keeping. Indeed, the "embryonic" development of clinical information systems constituted a fundamental barrier to the success of managed care organizations. In the 1990s, large health centers budgeted tens of millions of dollars per year to integrate systems that often were developed separately and almost never "talked to each other" (Shortell et al., 1994). A major challenge facing managed care is the design of information networks that provide direct clinical support to improve the process of care. However, despite the current enthusiasm for health information technology, its value

remains unclear. One recent analysis could not even find savings in hospital costs five years after the technology adoption (Agha, 2014).

A second feature of MCOs is their de-emphasis of the acute care hospital model. Hospitals provide expensive care, and moving toward cost-effective systems necessarily moves away from hospital care. As noted earlier, primary care physicians are often the gatekeepers of managed care systems, directing patients to appropriate (i.e., cost-effective) treatment settings. If they "feed" patients into the hospital instead, this leads to increased costs. Managed care seeks a vertical integration of what had previously been a generally unintegrated system of health care treatment. Through coordination of care and improved information, such integration has the potential to address the health care costs in a manner that would appear to address criteria of economic efficiency. Yet the integration is costly, and the quality of the resulting care may not match all consumer preferences. Some also claim that managed care systems have incentives to underprovide services, which may be harmful to patients.

The HMO represents a prime example of managed care on which there has been considerable research. We begin by describing HMOs and we continue with their history and with the rationale for a government policy that has promoted their development.

What Are the Economic Characteristics?

Managed care features a health care delivery structure involving the integration of insurers, payment mechanisms, and a host of providers, including physicians and hospitals. What distinguishes managed care from the fee-for-service care that also might plausibly attempt to integrate the various health care system parts?

Health insurance plans use four related mechanisms to contain costs and/or improve quality of care:

1 *Selective contracting*, in which payers negotiate prices and contract selectively with local providers such as physicians and hospitals. There may be price differences across providers and other contract features such as volume limits and discounts based on volume.
2 *Steering* of enrollees to the selected (in-network) providers. If patients select non-network providers, they may have to pay substantially higher out-of-pocket costs and, in some plans, pay the entire costs of these services.
3 *Quality assurance* through meeting voluntary accreditation standards. Practice guidelines, "best practices," and disease management programs are often incorporated into quality improvement activities.
4 *Utilization review* of the appropriateness of provider practices. The utilization review process may be prospective (in advance), concurrent (at the same time), or retrospective (looking back).

Of the four, most analysts find selective contracting to be most important. Dranove, Simon, and White (1998) point out that managed care organizations may award contracts on the basis of the providers' willingness to accept particular payment plans and monitoring of treatment styles and quality of care. Morrisey (2001) notes that under selective contracting (unlike FFS care), some providers get contracts and some do not. Service price becomes important in managed care system negotiations with providers.

The selective contracting and the steering distinguish managed care from the more standard FFS care. Managed care is also distinct through its quality assurance emphasis. Most plans seek NCQA accreditation. The NCQA, a private nonprofit organization, establishes performance

measures through HEDIS (Healthcare Effectiveness Data and Information Set), and it also issues report cards. Despite these efforts, as discussed later in this chapter, many analysts are disappointed at the progress managed care has made in improving quality. As for utilization review, Morrisey argues that traditional indemnity plans that include pre-hospital admission certification would be classified as managed care plans under this definition. Almost everyone now reviews utilization, so utilization review in itself is not helpful in discussing managed care.

What, then, is a good analytical way to conceptualize between MCOs and FFS? Cutler, McClellan, and Newhouse (2000) provide a useful conceptual model that asks how much a patient would have to be compensated to move from FFS to MCO coverage. The compensation presumably would be related to the patient's difference in utility (satisfaction) between FFS and MCO coverage. If an MCO and an FFS plan were identical, the compensation would be zero; if the MCO leads to less (more) utility, compensation must be positive (negative) to make the client indifferent.

Three differences between MCOs and FFS might affect compensation:

1 ***Difference in health.*** If the MCO provides reduced health (relative to FFS) due to reduced treatment, then the compensation must be positive for those who choose the MCO. This positive compensation might be offset if the MCO is better at managing the overall care process or at providing "well care."
2 ***Cost savings.*** If, holding health constant, the MCO provides savings due either to less treatment or cheaper treatment, the compensation must be negative, because the MCO is saving money for its clients.
3 ***Financial risk from different out-of-pocket payments.*** Clients may prefer an MCO if it ensures them from having to make large out-of-pocket payments. If so, the compensation will be negative because payment variability is reduced. The size of the compensation would depend on the MCO's cost-sharing provisions, as well as reimbursement for out-of-plan use.

This framework suggests that one must measure the differences between managed care and fee-for-service along several dimensions: health, price of care, and quality of care. In fact, patients who value health less (or other things more) may choose less health and/or health care by choosing an MCO, or possibly even no insurance, rather than FFS care.

It does not necessarily tell us which mechanism provides the appropriate level of care at which marginal benefits equal marginal costs. Recall that under FFS, with fractional coinsurance, clients may overuse services. Under managed care, they may use fewer services and possibly not enough of them, but it is not clear whether they will use the efficient amount.

With this framework established, we can look at the emergence of managed care plans and what the market for managed care will look like. We also can look at the differences in health, price, or quality of care, recognizing that consumers and employers, acting on their behalf, will evaluate all of these dimensions in spending their health insurance dollars.

The Emergence of Managed Care Plans

Managed care describes a variety of arrangements with the following common features. First, much, if not all, of the patient's care is provided through a specific network of hospitals, physicians, and other health care providers. Second, considerable centralized oversight of resource use, often referred to as utilization review, occurs within the network. Here, we provide a brief overview of the types of managed care plans and the extent to which physicians and hospitals now contract with insurers under capitation arrangements.

Employer-Sponsored Managed Care

Employer-sponsored insurance dominates the private health insurance market. In 2015, 147 million enrollees, nearly 55 percent of the nonelderly population, obtained coverage through the workplace. The Kaiser Family Foundation (KFF), through surveys and other sources, provides a wealth of information on the health care marketplace and employer-sponsored health plans.[1] The KFF data in Table 12.1 document the historic and dramatic shift to the three main types of employer-sponsored managed care plans. Traditional indemnity (FFS) insurance accounted for just 1 percent of enrollments in 2015 compared to 73 percent in 1988. HMO enrollments, which peaked at 31 percent in 1996 (not shown in Table 12.1), fell to 14 percent by 2015. PPOs, just 11 percent in 1988, represented 52 percent of enrollees in 2015, while point-of-service (POS) plans represented 10 percent.

Table 12.1 also shows the recent emergence of high-deductible health plans with a savings option (HDHPs). In 2015, these plans represented 24 percent of covered workers, up from 4 percent in 2006.[2] Much of this growth came at the expense of enrollments in the HMOs and PPOs.

Many employer-sponsored plans are partially or completely self-funded, i.e., self-insured. In contrast to fully-insured plans, where the entire risk is borne by the insurance company, an employer bears the financial risks under a self-funded plan, although it may purchase various insurance protections against unexpectedly large claims. Self-funded plans are governed by the Employee Retirement Income Security Act (ERISA) of 1974 which offers employers considerable flexibility by exempting them from state insurance laws including mandated benefits. According to the KFF, 63 percent of covered workers in 2015, and 83 percent of workers in firms with 200 or more employees, belonged to ERISA plans.

There are wide variations across health plans and provider organizations in terms of management of utilization and other features. With the emergence of many hybrid forms, the taxonomy of managed care is continuously evolving. For simplicity, we describe the three basic types of employer-sponsored managed care plans.

Health maintenance organizations (HMOs) provide relatively comprehensive health care, entail few out-of-pocket expenses, but generally require that all care be delivered through the plan's network and that the primary care physician authorize any services provided. Each subscriber is assigned a primary care physician ("gatekeeper") upon joining the HMO. If

Table 12.1 Health Plan Enrollment (Percent) for Covered Workers, by Plan Type, Selected Years

	1988	*1999*	*2008*	*2015*
FFS	73	10	2	1
HMOs	16	28	20	14
PPOs	11	38	58	52
POS	–	24	12	10
HDHPs	–	–	8	24

Source: Kaiser Family Foundation/Health Research and Educational Trust, *Employer Health Benefits, 2015 Annual Survey*, Exhibit 5.1. Available at www.kff.org/insurance: accessed January 15, 2016. This information was reprinted with permission from the Henry J. Kaiser Family Foundation. The Kaiser Family Foundation is a nonprofit private operating foundation, based in Menlo Park, California, dedicated to producing and communicating the best possible information, research, and analysis on health issues.

health care services are provided without gatekeeper authorization, then the HMO usually does not cover the services. The subscriber is personally liable for payment of the nonauthorized services. HMOs that directly employ physicians in their network are called staff model plans. In its simplest characterization, these physicians are paid salaries by the HMO, although some HMOs do base payments on factors such as patient load. Alternatively, plans that set up their network by contracting with physicians in geographically spread out, independent solo or small group practices are called independent practice associations (IPAs). Both types assign primary care physicians as gatekeepers for covered services. IPAs are more common than staff model HMOs.

Preferred provider organizations (PPOs) give subscribers two distinct tiers of coverage. When subscribers use the PPO's preferred provider network, the required cost sharing with deductibles or coinsurance is lower than when they use non-network providers. Although a network is formed, PPOs have no physician gatekeepers. Rather, patients simply must pay more out-of-pocket if they choose to go outside the network. In this way, PPOs create financial incentives for subscribers to use network providers rather than go outside the network for care.

PPO contracts with physicians and hospitals generally address the prices providers will charge the PPO. In return for promising to charge a lower price, selected providers become part of the PPO's preferred network. No guarantee is given that the provider will see patients under the plan, but if the network is not too large and the PPO's cost-sharing provisions for subscribers are network-favorable, then the provider may enjoy a large increase in patient care business by joining the network. Prompt payment for services is another advantage.

Providers often agree to submit themselves to some form of utilization review under the contract. Most PPOs require pre-admission certification for a hospital stay and concurrent utilization review for such stays. About half require a mandatory second opinion for a recommendation of surgery.

Point-of-service (POS) plans are a hybrid of HMOs and PPOs. Like PPOs, POS plans offer two tiers of insurance benefits. Coverage is greater (out-of-pocket costs are lower) when members use network providers and less generous (out-of-pocket costs are higher) when they use non-network providers. Like HMOs, however, POS plans assign each member a physician gatekeeper, who must authorize in-network care in order for the care to be covered on in-network terms. Most POS plans do not require authorization for a member to use out-of-network services, but such care is covered on less-generous terms.

Table 12.2 categorizes the organizational structures. The matrix rows indicate whether an organized provider network is formed. The columns indicate whether a gatekeeper is part of the arrangement. The gatekeeper and the provider network represent two particular forms of health system control. HMOs provide both, while FFS plans provide neither.

Table 12.2 Different Health System Organizational Structures

		Gatekeeper	
		No	*Yes*
Provider	**No**	Fee-for-service (FFS)	Point-of-service (POS)
Network	**Yes**	Preferred provider organization (PPO)	Health maintenance organization (HMO)

Medicaid and Medicare Managed Care Plans

In the last few years, many states have adopted managed care models for the Medicaid coverage they provide to families with dependent children and pregnant women who meet their low income criteria for Medicaid eligibility. They believe that managed care may help contain program costs, which are major parts of most states' budgets. As of 2013, nearly 45 million Medicaid beneficiaries nationwide (72 percent of Medicaid recipients) were enrolled in some form of managed care. This represents a sharp increase from the 2.7 million as recently as 1991. Tennessee had its entire Medicaid populations under managed care while others were planning to shift entirely to managed care.

As with employer plans, Medicaid managed care plans vary considerably. In some areas, states have contracted directly with HMOs that already exist in local markets. In others, states have created their own loosely structured provider networks, which contract with selected providers for discounted services and use physician gatekeeping to control utilization. Some Medicaid programs combine the two approaches.

Unlike Medicaid and private insurance, traditional fee-for-service coverage dominates Medicare, the federal program for the elderly. Of the nearly 55 million Medicare enrollees in 2015, 31 percent selected a Medicare Advantage plan (also known as a Part C plan). This was up from just 6 percent in 2005. Medicare Advantage plans are private plans that receive a fixed monthly amount per enrollee from Medicare. They include fee-for-service plans but HMOs and PPOs account for 88 percent of the enrollments.

The growth of Medicare managed care has resulted in major spillover effects to traditional Medicare and private insurance. Baicker and colleagues (2013) found that increases in Medicare Advantage enrollments, possibly through their effects on physician practice styles, produce substantial system-wide reductions in hospital stays and hospital costs.

There are significant ongoing developments that are relevant to both Medicaid and Medicare managed care, and the major ones are further discussed later in this chapter as well as in in Chapters 20 (Social Insurance) and 22 (Health System Reform). We note that Medicaid managed care enrollments in those states that have expanded their programs under the ACA[3] received a major boost. Starting in 2014, Medicaid eligibility included all individuals under 65 with incomes below 133 percent of the federal poverty level. This contrasts with Medicaid's historic emphasis on coverage for pregnant women, children, and the disabled.

Managed Care Contracts with Physicians

Managed care contracts with physicians vary considerably. Most HMO and POS plans pay their network physicians on a capitation basis. Under capitation, the plan pays the physician's practice a fixed fee, generally an actuarial per-member-per-month (PMPM) dollar amount, in return for the treatments provided to members of the insurance plan. Physicians also may be responsible for the costs of referrals, laboratory tests, and hospital services. Thus, HMO and POS plans shift the costs of care, as well as the risk associated with those costs, directly onto physician practices. In so doing, these contracts put physician earnings at risk. If care provided under such arrangements turns out to cost less than the fixed-dollar plan payment, the practice makes a profit. If instead care costs more than the payment, the practice must take a loss.

In contrast, PPO contracts with physicians rarely involve capitation. Instead, they specify the discounted fees for various services that the plan will pay in exchange for the privilege of being in that plan's network. If a physician joins the PPO's network and happens to provide services

to one of that plan's subscribers, the practice must accept the pre-negotiated fees as payment in full. "Balance billing" of the patient (for the remainder of a higher bill) is not allowed.

Managed care contracts, whether they are HMO, PPO, or POS plans, commonly contain utilization review procedures for physicians. Most managed care contracts also require a certain degree of physician record-keeping on their enrollees (e.g., plan-specific patient encounter forms may have to be filed with the insurer each time care is provided).

Medicaid managed care contracts with physicians parallel those of private managed care plans, although specific service packages are determined heavily by the state's policies. In states that have set up their own Medicaid provider networks, the state contracts directly with individual gatekeeper physicians, agreeing to pay them a small fixed fee for each Medicaid enrollee under their jurisdiction. In return for this payment, the physician serves as the gatekeeper for Medicaid-covered services.

Managed care contracts are nearly universal in physician practices. In 2008, 88 percent of physicians had managed care contracts with 70 percent having five or more contracts (Boukas, Cassil, and O'Malley, 2009). Capitation arrangements were once seen as important to cost-containment efforts but their role has declined sharply over the past 20 years. Zuvekas and Cohen (2016) reported that only 18 percent of office-based physician visits by private HMO patients, and 8 percent of Medicaid visits, in 2013 were reimbursed under capitation arrangements. Despite concerns about the adverse incentives created by fee-for-service, it accounted for 95 percent of the reimbursement for all visits.

Managed Care Contracts with Hospitals

HMO and PPO plans contract with only a subset of the providers (physicians and hospitals) in the areas that they serve. This key feature of the managed care sector allows plans to promote price competition among hospitals that might otherwise lose plan business.

As recently as the early 1980s, fewer than 20 percent of the insured population was enrolled in managed care plans, with most in Kaiser-model HMOs (named after the Kaiser-Permanente HMO system) where the HMO owned the hospitals that its members used. In the 1980s, many states passed "selective contracting" laws, which provided insurers with greater flexibility to develop alternative health plans and to test different design features. These laws led to growth in PPOs and allowed more flexibility than the "closed-system" HMOs such as Kaiser.

By 2005, the KFF estimated that the proportion of hospitals reporting revenue from capitated contracts had increased to 38 percent (from 30 percent in 1998). Hospitals in urban areas, and particularly inner-city facilities, are more likely to report capitation revenues than are rural hospitals.

The probability and characteristics of contracts between individual managed care organizations and hospitals appear to depend on three sets of factors:

1 *Plan characteristics*, including whether it was a PPO or an HMO (and possibly what type of HMO), plan size, whether the plan serves several localities, and how old the plan is.
2 *Hospital characteristics*, including size, ownership (including for-profit versus nonprofit status), location (city versus suburb), teaching status, and cost structure (reflecting prices).
3 *Market characteristics*, generally measured at the metropolitan area level, including the penetration and rate of growth of managed care plans.

Research has found equivocal results, most often on the important issue of hospital costs, which are used to reflect prices to the plans. Early studies found that before managed care

plans became popular, more competitive markets had higher hospital costs. This occurred because under cost-based FFS reimbursement, hospitals could (and did) compete on the basis of services and quality rather than price. More recent research has suggested that competition in hospital markets can lead to lower costs when the insurance market includes sufficient managed care penetration.

Development and Growth of Managed Care—Why Did It Take So Long?

Why did it take managed care so long to become a force in the marketplace? A variety of institutional, economic, and political forces has influenced its development and growth. Above all, the story of managed care in the United States requires an account of the strenuous historical opposition from organized medicine. Early on, organized medicine fiercely opposed alternatives to free patient choice and particularly alternatives to FFS reimbursement. These alternatives posed a threat to a physician's ability to earn excess profits (known as economic profits or rents). Kessel (1958) described the historic political actions of organized medicine and hypothesized that FFS enabled physicians to charge some patients a higher fee than others for essentially the same service, a pattern known as price discrimination. Such price discrimination is difficult or impossible under the contracts that characterize prepayment-based organizations. First, providers will find it difficult to determine how much individual consumers value the services. Second, the prepayment-based organization may be able to shop among providers, thus limiting the providers' monopoly power.

Organized medicine also created other barriers to managed care development. For example, it opposed physicians' participation in plans that were not controlled by physicians and/or that were not offering a free choice of physician. The Federal Trade Commission successfully challenged such restrictions in the 1970s.

Federal Policy and the Growth of Managed Care

The HMO Act of 1973 represented a turning point in federal policy in promoting the development of alternative delivery systems as a cornerstone of a cost-containment strategy. The act enabled HMOs to become federally qualified if they provided enrollees with comprehensive benefits and met various other requirements. Loan guarantees and grants for startup costs were made available, but the main advantage accruing to a federally qualified HMO was that it could require firms in its area with 25 or more employees to offer the HMO as an option. Other regulatory barriers subsequently were reduced.

Despite these changes, the number of HMOs and HMO enrollees did not accelerate until the 1980s. When incentives to enroll Medicare and Medicaid recipients improved, the entry of for-profit HMOs led to growth from 235 in 1980 to 623 in 1986. Growth slowed in the late 1980s and some consolidation occurred in the number of HMOs, but the upward trend in enrollment resumed in the 1990s. While the number of plans declined slightly from 1990 to 2000, total enrollment increased from 33 million to 81 million persons. Since then, HMO enrollments and the number of plans have decreased, but the gap has been filled by the growth of PPOs, which offer more flexibility in choosing providers. For example, HMO and PPO memberships in employer-sponsored plans were roughly equal in 1996 at 31 percent and 28 percent, respectively, of covered workers. As seen in Table 12.1, 52 percent chose a PPO in 2015, compared to just 14 percent choosing a HMO.

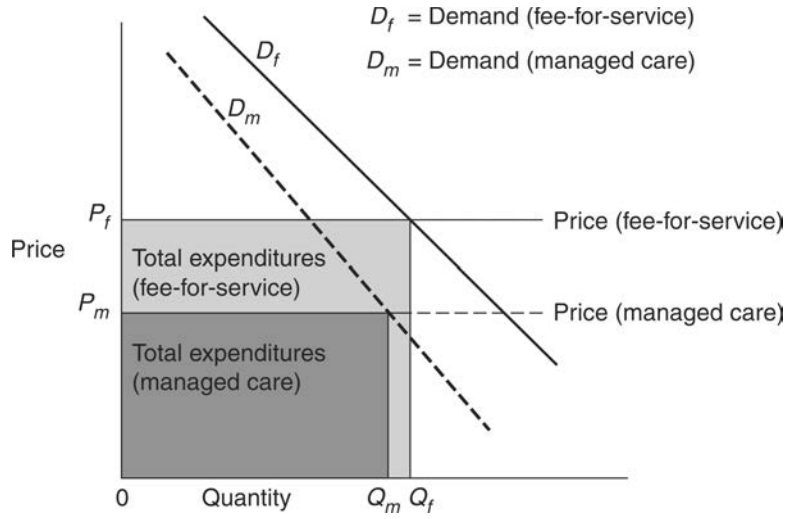


Figure 12.1 Treatment and Expenditures under Managed Care

The Economics of Managed Care

The erosion of barriers to prepaid plans, coupled with some provider markets characterized by economic profits, gives us the key elements to understand the rationale for managed care. Figure 12.1 starts with consumer demand D_f under fee-for-service insurance. Assuming that the price is constant at P_f, total expenditures, $P_f Q_f$, are represented by the larger rectangle. These total expenditures also would represent the expected insurance component of the individual's wage package.

A managed care option by constraining choice of provider as well as various coverages is viable only if it reduces expenditures. By exerting market power over suppliers, HMO managers may lower prices from P_f to P_m. A price reduction is possible when providers have been earning economic profits.[4] In effect, managed care reduces or eliminates those economic profits.

Demand also may be reduced from D_f to D_m by reducing inpatient care, by limiting length of stay, minimizing supplier-induced demand, and, in general, by encouraging more cost-effective care through the use of information technology and financial incentives to providers. The resulting expenditures are shown by the smaller rectangle representing total expenditures, $P_m Q_m$.

It follows that managed care trades some constraint of consumer choices for lower per-unit prices for care. As drawn in Figure 12.1, both decreased price per unit of care and decreased quantity of care contribute to decreased expenditures. Note, however, that a natural response to decreased prices is to increase quantity demanded. Total consumer expenditures will fall as long as the price decreases are not fully offset by increased quantity demanded.

Modeling Managed Care

The changed incentives from managed care have caused many to predict success in constraining utilization and cost of care. The underlying logic behind this proposition is

intuitive. When providers agree to handle all of their patients' health care needs for fixed, prearranged fees, the providers accept and bear a substantial part of the financial risk. By bearing the risk of delivering services at a fixed premium, managed care plans have strong incentives to reduce excessive care. The fixed premiums would also seem to provide incentives to offer forms of preventive care that are cost-effective from the care managers' perspectives. Such preventive care could decrease the need for more subsequent (expensive) curative care.

Will managed care plans provide enough care or the right types of care? The following analyses, developed by Goodman and Stano (2000), treat managed care plans as "staff model" HMOs where physicians are HMO employees.[5]

Modeling Individual HMOs

Individual HMOs need to determine the number of consumers to serve, or quantity, and the level of service to provide, or quality. One might compare them to urban suburbs, which set the qualities of public schools and the sizes of the police forces, for example, by the preferences of their residents and by the costs of providing them. Just as suburbs provide menus of public services, HMOs provide menus of care in inpatient and outpatient settings. To keep things simple, we will assume that HMOs provide only one type of service (visits), and that the HMOs are differentiated in quality by how many visits each offers.

We assume that HMO treatment costs are related to member health status, which is a function of care received at the HMO and elsewhere, as well as other factors including diet, environment, and lifestyle. People's long-term health relates in part to short-term decisions about how much care to provide. This follows directly from the long-term nature of human capital, most particularly health capital. Thus, because patients live for many years, treatment decisions at one HMO may affect treatment costs at other HMOs. Moreover, HMOs have the dual incentives of keeping people healthy and attracting healthy people.

An HMO's total annual costs are higher if it provides more services per enrollee (quality) or if it has more members (quantity). Having healthier members lowers HMO costs. Assume that health care works—that health is related positively to the level of services by all providers. Because at any time in the future these individuals may be members of a given HMO, treatment of the whole population by all providers affects the average health status of the population. This will influence the costs for these providers in that better health will lower costs.

How Much Care?

The previous section indicated that profit-maximizing HMOs might not recognize the system-wide impact of health services on consumers' health. Although an individual HMO seeks to maintain the health of its own clientele, it does not recognize that its decisions may affect the costs of other HMOs. We call this an externality because it is a benefit that affects others but is not considered by any individual HMO.

Suppose instead that an HMO has clinics at a number of locations, referred to as HMO_1, HMO_2, and so on. Although HMO_1, for example, may not account for this possibility, the HMO entrepreneur, to maximize profits, must consider the impacts of health services on the health of others elsewhere.

We see this in Figure 12.2. Without the externality, HMO_1 optimizes at point A, providing quality level x_1^{mkt}. Level x_1^{mkt} is economically inefficient, however, because it does not account for the fact that improved treatment at HMO_1 lowers costs throughout the system.

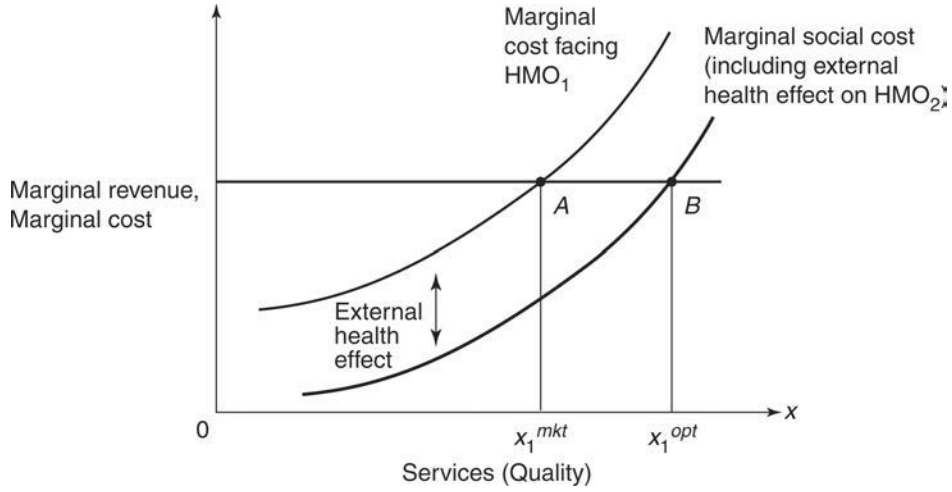


Figure 12.2 Externality Model of HMO

The optimal quality level of x_1 is at point B, or x_1^{opt}. This occurs by recognizing the downward shift in the marginal cost curve by a factor that reflects the effect of the health externality on the costs facing HMO_2, HMO_3, and so on, as well as HMO_1. Level x_1^{mkt} indicates an inefficiently small level of HMO services, and by implication a substitution of non-HMO and/or nonhealth care inputs (such as the patient's own time) for the HMO care.

Although the potential loss of patients may influence treatment decisions of FFS providers as well as HMOs, the capitation method of payment to HMOs makes the problem of potential disenrollment particularly important. Fee-for-service providers are paid for each unit of care. Aside from uncollectible bills, they do not risk losses on services provided currently or in the future.

In contrast, by integrating insurance with the provision of health care, the HMO receives a fixed payment per enrollee to cover costs in the current period, and over time, for those who remain enrolled. Unlike FFS care, where payment in every period is likely to cover costs, the HMO must consider the timing of expenditures and the financial losses of overspending on patients who may disenroll. One way for an HMO to protect against long-term losses attributable to disenrollment is to economize on care for those currently enrolled.

What Types of Care?

In this section, we consider the long-term consequences of potential movement among HMOs on decisions among treatment practices. If an HMO receives constant revenue per patient each period, its goal is to minimize costs. Patients stay in the HMO for two periods, which might be considered as early and late in their lives. The HMO may offer:

- High-tech, possibly capital-intensive procedures leading to high Period 1 costs and zero Period 2 costs.
- Low-tech, less capital-intensive procedures, leading to low costs in both periods.

Because the revenues are the same with either high-tech or low-tech care, the HMO's problem is to minimize costs, using the cheaper (over time) of the two procedures.

Suppose that increased competition through increased choice raises disenrollment rates. With higher disenrollment rates, low-tech continuing care becomes the more financially viable option, even if the present discounted values are equal and even if high-tech treatment is more economically efficient in producing health. Here, HMOs protect themselves against future disenrollment by reducing current costs through (low-cost) continuing care rather than high-tech treatment.

In evaluating how much, and what types of, care HMOs offer, we see that the HMO faces an economic externality because it cannot capture fully the gains of its treatment over time. As a result, it will offer less care and lower-tech care than FFS plans. As noted, if the HMO were to merge with others forming a larger network, the larger firm might internalize this positive externality because the network owners could expect that clients who leave one HMO might join another HMO within the network. The receiving HMO would then take advantage of the now-healthier clients who had benefitted from their earlier treatment.

Framework for Prediction

Clearly, this model simplifies the situation because many forms of managed care exist providing a myriad of services. It provides a framework, however, for addressing possible HMO cost savings relative to FFS plans. FFS plans encourage overutilization to the point where marginal private benefits can be far less than marginal costs. HMOs are widely believed to discourage this deadweight loss and other forms of overutilization, such as supplier-induced demand. To evaluate utilization, however, one must control for the health of the client population, which might be impacted by HMO "cream-skimming" of the healthier clientele.

This simple model provides predictions that are consistent with the rising conflict between HMOs and their members. The popular press reports stories about inadequate levels of services provided by HMOs, and the unavailability of expensive, high-tech treatment options.

Where Managed Care Differs from FFS—Dumping, Creaming, and Skimping

We recognize that one of the key differences between HMOs and FFS plans involves the form of payment. Hospitals and other FFS providers are paid for each treatment in order to cover costs. HMOs are paid fixed rates per person irrespective of the amount of treatment used. These differences have led to discussions within the medical and policy communities regarding three purported practices.

1 ***Dumping.*** Refusing to treat less healthy patients who might use services in excess of their premiums.
2 ***Creaming.*** Seeking to attract more healthy patients who will use services costing less than their premiums.
3 ***Skimping.*** Providing less than the optimal quantity of services for any given condition in a given time period.

Because the HMO's costs will depend on the average health of its clientele, practices such as dumping (unhealthy patients) and creaming (healthy patients) may occur. If the HMO can identify and dump patients who are sicker, it can improve the average health of its clientele and hence lower its costs. As noted in Ellis (1998), the patient and the public at large may not notice this decision. The HMO may honestly say that it is not equipped for this kind of case and that the patient would be better served elsewhere. This behavior contrasts with cost-based FFS care. Inasmuch as every penny spent on even the most severe case theoretically is paid back to the hospital, the cost-based hospital will not dump anyone.

Creaming is the practice of seeking out or emphasizing low-severity patients. The HMO benefits from creaming because all patients of sufficiently low-case severity require few hospital services, so that premiums for these patients exceed their costs. Some patients believe that their malady is so mild that the hospital care is hardly worth the trip; these patients correspond in practice to the young, vigorous, and healthy people that HMOs prefer to have in their service populations. The FFS provider also creams because its reimbursement covers costs for each of its patients.

Skimping involves cutting back on services to the point that patients' welfare is reduced. FFS providers will not skimp because each nursing hour, electronic instrument, and surgery that they employ will be reimbursed. In contrast, skimping provides the opportunity for the HMO to increase profits. Because the HMO has received the premium in advance, reducing the amount of services will lower costs and may raise profits. The uncertainty regarding profits occurs because reducing the benefits for a given severity of illness may cause the HMO to lose some patients to FFS plans or to other HMOs.

Equilibrium and Adverse Selection in a Market with HMOs

Providers may have incentives for seeking patients, but patients also may select themselves into particular types of care. Cutler and Reber (1998) demonstrate the potential adverse selection of sicker consumers toward FFS care and healthier patients toward HMOs. This work also explains how HMO penetration can respond strongly to small changes in relative pricing.

Suppose that Jeff has just graduated from college and landed a job with Santa Fe Futons. Santa Fe offers its employees a choice of membership in one of two health plans. The HMO charges a flat prearranged price P_{HMO}; the deluxe care FFS plan features deductible D and coinsurance rate r. Letting the severity of the illness, s, be measured by the expenditures required to treat at the FFS level, Jeff calculates the extra cost E of FFS to be:

$$E = \text{FFS cost} - \text{HMO cost} = (\text{Deductibles} + \text{FFS copay}) - \text{HMO price}$$
$$E = (D + rs) - P_{HMO} \tag{12.1}$$

Is the extra cost, E, to use FFS worth it, or should Jeff use HMO care instead?

To evaluate Jeff's decision, we calculate market values for the terms in Equation (12.1) and then compare term E to the value received from FFS care. If this extra cost of using FFS exceeds the value to Jeff of FFS, he will choose HMO care. If not, he will choose FFS. Under perfect competition, both FFS and HMO providers earn zero profits in the long run. FFS provider profit π_F for the average patient is written as:

$$\pi_F = D + \overline{rs} - \overline{s}_F = 0$$

indicating that the firm collects patient deductibles D and patient copayments $\overline{rs}$, and spends $\overline{s}_F$ for treatment (where the "bars" indicate FFS and HMO averages). Rearranging terms, this means that the market deductible is:

$$D = (1 - \overline{r})\overline{s}_F$$

For the HMO parameters, we again assume that in the long run, competition will drive profits to zero. Cutler and Reber assume that HMOs achieve efficiencies in providing care, as

well as restricting the amount of hospital care provided, reducing the cost of care. They summarize the efficiencies using parameter α. With no cost reductions, α would equal 1; Cutler and Reber assert that the real-world value of α is about 0.9. As a result, premium P_{HMO} must cover average payments for condition $\overline{s}_H$, or $\alpha\overline{s}_H$, so that:

$$\pi_H = P_{HMO} - \alpha\overline{s}_H = 0$$

Substituting these terms into (12.1) provides the following expression for Jeff:

$$E = [(1 - \overline{r})\overline{s}_F - \alpha\overline{s}_H] + rs \tag{12.2}$$

Thus, E consists of a market-determined constant (in brackets) plus an increment of expenditures, rs, depending on the severity of the condition. It is reasonable to assume that the market-determined term is positive, so that E is the upward-sloping line in Figure 12.3.

To this point, we have looked at extra costs of FFS care. We also recognize that Jeff, like many, may see an additional benefit in FFS, the benefit of being able to choose one's own doctor. Cutler and Reber believe that this additional benefit would increase with the seriousness of the illness, in that people with serious illnesses would prefer to choose their own physicians. If so, the additional benefit of FFS starts at the origin (if $s = 0$, there is no benefit) and rises as s increases. We label this curve V in Figure 12.3.

Start at expected level of care $s = 0$ where Jeff is young, healthy, and does not expect to use much service. If he were to join the FFS plan, he would have to pay the average FFS client expenditure, which exceeds the average HMO fee (the term in brackets in equation 12.2), plus the share r of the services s that he uses. Because the extra FFS costs are higher than the value V that he puts on them (the E curve is above the V curve), Jeff chooses to join the

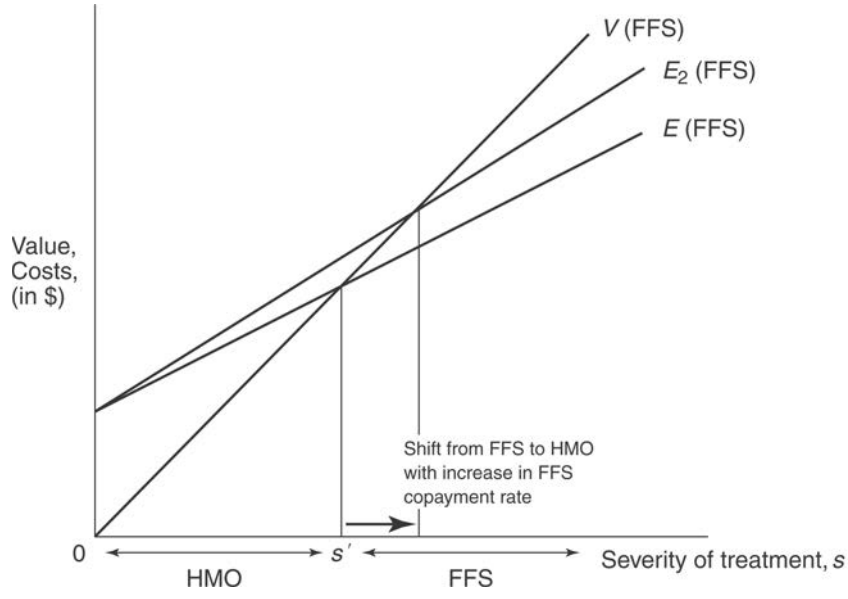


Figure 12.3 Selection into HMO and FFS Settings

HMO. If Jeff has a chronic condition requiring ongoing treatment, he recognizes that each unit of ongoing treatment is subsidized by the FFS plan at the rate $(1 - r)$. With increasing severity of treatment, s, FFS becomes a more attractive option as V exceeds E.

At severity level s', Jeff would be indifferent between the HMO and the FFS plans because the extra value of FFS just equals the extra cost. If all consumers were similar to Jeff in everything except their health, consumers with expected severity less than s' would join the HMO, while those expecting severity greater than s' would choose FFS.

Suppose the FFS plan increased its coinsurance rate r. Then the FFS plan becomes less attractive; this can be seen by rotating the E line counterclockwise to E_2. Consumers who were previously indifferent between HMO and FFS plans will shift to the HMOs. The younger and healthier HMO patrons are now joined by some of the (not as young and not as healthy) former FFS patrons. As a result, the healthiest among the former FFS patrons become the sickest HMO members and the average severity of illness in both plans increases.

How Does Managed Care Differ?—Empirical Results

Economic and organizational theories have suggested that managed care will differ from more traditional fee-for-service plans. One might predict that managed care organizations will spend less per member, reducing health care costs. Theory would also predict, however, that if fewer resources are used, quality of care may also suffer. Policymakers have considerable interest in whether this theoretical proposition is true. Researchers early on reported that total costs—that is, the sum of premiums and out-of-pocket expenses—were from 10 to 40 percent lower for HMOs. They attributed the cost differences largely to lower hospitalization rates, not to lower ambulatory care-use rates.

Methodological Issues—Selection Bias and Quality of Care

Conceptually, it would seem fairly simple to compare health care costs in managed care and fee-for-service plans. Researchers would collect data on cost of care across a wide spectrum of the population. Controlling for items such as patient age (older people have higher costs) and existing health status (sicker people have higher costs), one could use multiple regression statistical methods to compare costs.

Two major issues complicate the comparison—selection bias and quality of care. The methods previously discussed work only if patients are randomly assigned to either HMO or FFS treatment. Analysts worry that this random assignment does not exist in the real world. On the one hand, HMOs offer comprehensive benefits and so they may attract and retain sicker members. If we do not address this feature, studies may make HMOs look more expensive than they really are. On the other hand, HMOs may attract disproportionately younger members and families who tend to be healthier, and for whom the costs of care tend to be relatively lower. Studies that ignore this problem may make HMOs look less expensive.

Does managed care offer quality of care that is comparable to care under FFS? Managed care provides incentives to reduce the costs of care. Does it also provide incentives to cut corners by reducing the quality of care? Although some consumers may choose to pay less for lower quality care (just as some buy cheaper tires or cheaper cuts of meat), it is essential both to measure quality and to control for quality differentials in evaluating differences in health care costs.

The definition of quality is by no means obvious. Cutting health care quality would likely lower costs in the short term, but it might increase the longer-term costs if patients required

additional services later. Furthermore, if information about quality were available to consumers, lowering quality would tend to erode demand. Despite the importance of quality-of-care issues to health care analysis in general and to analysis of managed care in particular, quality issues are difficult to resolve.

Quality may range from consumer perceptions of the provider–patient relationship to the outcome effects of health care on health status. Donabedian (1980) provides three general descriptors:

1 ***Structure.*** The quality and appropriateness of the available inputs and their organization.
2 ***Process.*** The quality of the delivery of care.
3 ***Outcome.*** The ultimate quality of care but the most difficult to measure scientifically.

Comparative Utilization and Costs

Due to the evolving nature of both managed care and fee-for-service provision, some of the previous studies are more interesting for their historical perspective than for their current applicability. Luft (1978, 1981) found that HMO enrollees, especially prepaid group practice members, had lower hospitalization rates. No clear evidence showed that these lower rates were attributable to reductions in the less important, discretionary procedures. Furthermore, the evidence at hand could not dismiss the possibilities that biased self-selection of HMO membership or underutilization in HMOs was responsible for the observed differences.

Arnould and colleagues (1984) confirmed Luft's conclusion that length of stay was not significantly different between the HMO and the FFS patients. They also found that the use of surgeon visits, as well as lab charges, per patient were lower for the HMO users (significantly lower for hysterectomy and appendectomy), although total hospital charges were significantly lower for the HMO patients only in the case of appendectomies. Thus, although differences occurred in costs of elements of hospital care, no strong case could be made to conclude that HMOs produce hospital care more cheaply overall.

The RAND Study—A Randomized Experiment

In the RAND Health Insurance Experiment, patients were assigned randomly to different plans in a controlled experiment, thus apparently eliminating selection bias. Would HMO costs still be lower under such circumstances?

The RAND study (Manning et al., 1984) compared HMO and FFS patients in the Puget Sound area, where 1,580 individuals were assigned randomly to either an FFS physician of their choice or to the Group Health Cooperative (GHC) of Puget Sound—an HMO in Seattle, Washington. The 431 FFS individuals were in one of four groups:

1 Free care.
2 25 percent of expenses up to a maximum out-of-pocket liability of $1,000 per family.
3 95 percent of expenses up to a maximum out-of-pocket liability of $1,000 per family.
4 95 percent coinsurance on outpatient services, up to a limit of $150 per person ($450 per family).

In addition to the experimental GHC group of 1,149 persons, a control group consisted of a random sample of 733 GHC members who had been enrolled for at least one year.

Total expenditures per person were $439 (in 1983 dollars)[6] for the experimental group, including out-of-plan use (which may be substantial), compared to $609 for the free care FFS group (group 1). Ambulatory utilization was about the same. Thus, the 39 percent increased

spending for FFS members (or 28 percent reduction for GHC) was due largely to a much higher admission rate and increased hospital days per person. The study could not pinpoint the reasons for GHC's lower hospital use.

To put the potential cost savings into better perspective, use rates for the experimental HMO patients did not differ materially from those of FFS groups 3 and 4 above. Thus, for some population groups, a shift to HMOs would not lead to savings, although the cost savings for other population groups might conceivably be important.

More Recent Evidence

In a series of studies, Miller and Luft (1994, 1997, 2002) summarized findings regarding quality of care, utilization, and customer satisfaction. In the late 1980s and early 1990s, managed care (compared to FFS) plan enrollees received more preventive tests, procedures, and examinations (such as cancer screening; pelvic, rectal, and general physical examinations). Outcomes on a wide range of conditions (including congestive heart failure, colorectal cancer, diabetes, and hypertension) were better or equivalent to those using FFS plans. HMO enrollees were less satisfied with quality of care and physician–patient interactions but more satisfied with costs.

Their 1997 article found that HMO plans and providers cut hospitalization and use of more costly tests and procedures, often with little visible effect on quality of care "given the high prices of the indemnity insurance/fee-for-service system." However, simply carrying out the same clinical processes but with fewer resources can negatively affect quality of care in some cases, such as Medicare home health care.

In their 2002 review, which covered the period 1997–2001, Miller and Luft included HMOs and some mixed models, but excluded studies purely on PPOs. Table 12.3 summarizes their analysis along six dimensions: (1) quality of care, (2) access to care, (3) satisfaction, (4) prevention, (5) length of stay, and (6) use of expensive resources. In the first line, for example, 14 studies showed better (HMO) quality, 15 showed worse quality, and 18 were in the middle. The overall results summarized in Table 12.3 are similar to those in the two previous articles. Compared with non-HMOs, HMOs had similar quality of care, more prevention activities, less use of hospital days and other expensive resources, and lower access and satisfaction ratings.

Table 12.3 HMO Plan Performance Update: An Analysis of Published Literature, 1997–2001

Indicator	*Favorable to HMOs*	*Mixed*	*Unfavorable to HMOs*	*Total*
Quality of care	14	18	15	47
Access to care	2	4	4	10
Satisfaction	0	3	8	11
Prevention	7	3	0	10
Length of stay	5	5	0	10
Expensive resources	8	7	0	15

Source: Miller and Luft (2002).

BOX 12.1

What Do HMOs Actually Do?

There seems little doubt that HMO expenditures per member are substantially lower—sometimes 30 to 40 percent lower—than expenditures in traditional indemnity plans. Are these reductions achieved mainly by managing access and utilization, and lowering payments to providers that reduce their economic profits? What is the role of risk selection and quality? Several contributions provide considerable insight into these challenging questions.

Using methods described earlier in this chapter to decompose differences between FFS and HMOs, Cutler, McClellan, and Newhouse (2000) focused on Massachusetts patients with newly diagnosed heart disease, both those with heart attacks (acute myocardial infarction), which are relatively expensive to treat, and those with less severe forms of ischemic heart disease. By selecting one condition, the authors avoid some of the problems associated with aggregation across conditions. By studying heart disease, risk selection is minimized because even if a patient chose a plan based on some expectation of heart disease, the choice would not likely be based on expectations regarding the severity of the disease.

The study results are very clear and powerful: "Essentially all of the differences in reimbursement . . . [result from] differences in the prices paid for particular services, rather than differences in quantity or quality of services received" (p. 327). The authors caution about generalizing findings based on a life-threatening condition, such as a heart attack, where insurance status may have little effect on treatment. They also suggest that cardiac care is well known for providing sizable economic rents to both hospitals and cardiologists. Thus, the price effect may not be as large for other kinds of services.

Polsky and Nicholson (2004) also decompose the differences between HMOs and non-HMOs into differences in risk selection, utilization, and prices. They use a national sample for overall expenditures which were $188 (9.3 percent) lower for HMO members. Consistent with the Cutler study, lower prices paid by HMOs were the main determinant of expenditure differences. Prices were actually $269 less for the HMOs, or more than the expenditure difference because their utilization was $81 higher. Risk selection accounted for only $35 of this amount.

With their consistent results, the two studies provide a clearer picture of the role of HMOs. HMOs are quite successful in using their leverage to negotiate lower fees, and they do this without obvious reductions in quantity or quality. Risk selection also is not a major factor. However, it must come as a disappointment to many proponents of managed care that there is no indication that HMOs have been able to fulfill their promise and potential of applying information technology and better management to improve the process of health care delivery.*

Note: *We do not mean to imply that MCOs have not engaged in major effort to improve quality, but, until recently (Landon et al., 2008), this effort has not been systematically described. Wu (2009) also provides new information about hospital price discounts. Large plans and those with a greater ability to channel patients to alternative hospitals are able to extract higher discounts

Other work confirms some of these findings (see also Box 12.1). Rizzo (2005) concludes that HMO patients get substantially more preventive care than FFS patients and that this is not due to a selection effect that patients/physicians with preferences for preventive care are more likely to choose HMOs. After accounting for self-selection, Deb and colleagues

(2006) estimate that an individual in a managed care plan would receive about two more physician visits and 0.1 emergency room visits per year than had the same person enrolled in a non-managed care plan. Because of these and other research developments, we now have a much better understanding of the performance and role of HMOs.[7]

Growth in Spending

Analysts believe that managed care reduces utilization, especially of hospital care. A different but related question is whether managed care organizations also have lower growth rates in spending. If they do, a continued shift toward managed care will result not only in reductions in spending levels, but also in the long-term rate of increase.

It is important to provide a framework for discussing the relationship among FFS, MCOs, and total costs, particularly because terminology can be confusing. Suppose we are concerned about costs per person for treating a particular illness over three periods. Let us assume that people use either FFS or MCO and that the population is fixed. We can calculate the total treatment costs as:

$$\text{Total treatment costs} = (\text{Number in FFS}) \times (\text{FFS costs/FFS enrollee}) + (\text{Number in MCO}) \times (\text{MCO costs/MCO enrollee}) \quad (12.2)$$

Dividing both sides by the total population, we get:

$$\text{Treatment costs/Person} = (\%\ \text{of population in FFS}) \times (\text{FFS costs/FFS enrollee}) + (\%\ \text{of population in MCO}) \times (\text{MCO costs/MCO enrollee}) \quad (12.3)$$

Suppose, in Period 1, that FFS treatment costs $2,000, MCO treatment costs $1,000, and that 60 percent use FFS and that 40 percent use MCO care. The treatment costs per person will be:

$$\text{Treatment costs/Person} = (0.6 \times \$2{,}000) + (0.4 \times \$1{,}000) = \$1{,}200 + \$400 = \$1{,}600$$

This is in column 1 of Table 12.4 as Period 1.

If FFS and MCO costs were to stay constant and patients were to switch from the more expensive FFS to the less expensive MCO, costs per person would fall. As noted in Table 12.4, if a 10 percentage-point movement occurs from FFS to MCO, treatment costs per person would fall because 1 in 10 people would be substituting a (MCO) treatment that is half the price of the other (FFS treatment). As calculated in Table 12.4, the total costs per person would fall by $100, or 6.25 percent.

What is less obvious is that unless the population shift from FFS to managed care continues, cost reduction and cost containment may be difficult. Suppose that in Period 3 the percentage in FFS remains at 50 percent, but the costs of both FFS and MCO increase by 10 percent. The total cost per person accordingly increases by 10 percent from $1,500 to $1,650. If total costs per enrollee in each sector were to continue to increase by 10 percent and no change occurred in the MCO market share, the total costs per person on aggregate also would increase by 10 percent. This occurs irrespective of the fact that a larger proportion of the population (Period 3) is now being treated in managed care settings than was being treated in Period 1.

Table 12.4 Managed Care and Cost Containment—An Example

	Period 1	*Period 2*	*Percent Increase*	*Period 3*	*Percent Increase*
Fraction of population MCO	0.4	0.5		0.5	
FFS costs per enrollee	2,000	2,000	0.00	2,200	10.00
MCO costs per enrollee	1,000	1,000	0.00	1,100	10.00
Total costs per person	1,600	1,500	–6.25	1,650	10.00

This analysis provides important insights into the impacts of managed care plans on health care costs and health care cost increases. To the extent that large shifts of insureds into managed care have led to lower-cost treatments, there may have been one-time cost decreases relative to what they would have been. However, if managed care and FFS plans face the same cost inflation for the services that they offer apart from the one-time decreases, overall cost inflation is unlikely to abate.

Early studies by Luft (1981) and by Newhouse and colleagues (1985) found the growth rate of HMO spending to be roughly the same as the growth rate under FFS, and recent studies have not contradicted those findings. In the 1990s, policymakers relied on shifts into managed care to reduce health care costs and their growth rates. With most persons outside the Medicare program now covered by a managed care plan, such shifts can no longer continue. In the absence of other innovations, costs will continue to inflate, albeit from lower bases.

Competitive Effects

Up to this point, we have concentrated on the direct effects of managed care and managed care organizations. We have asked what managed care organizations look like, whether they provide less costly care, and whether they provide different quality care. In this section, we address the indirect effects of managed care through the market mechanism. By indirect, we mean that existing health providers must respond to competition from the managed care sector.

After discussing some of the theoretical concerns, we will evaluate the impact of competition in three ways. We will look at the impact of managed care on hospital markets, on insurance markets, and finally on the adoption of technological change.[8]

Theoretical Issues

The spread of alternative delivery systems can elicit substantially greater competition in other sectors only if competition is absent at the start. Otherwise, both providers and insurers would be operating at, or close to, their minimum costs of production. Although there is bound to be disagreement on the extent of the degree of market imperfections, most would agree that the insurance and provider markets are less than highly competitive.

Consider the consequences of having a larger number of products and competitors to an existing monopolistic seller. In Figure 12.4, D_1 is the current market demand and P_1 is the monopoly profit-maximizing price (where marginal revenue MR_1 equals marginal cost MC)

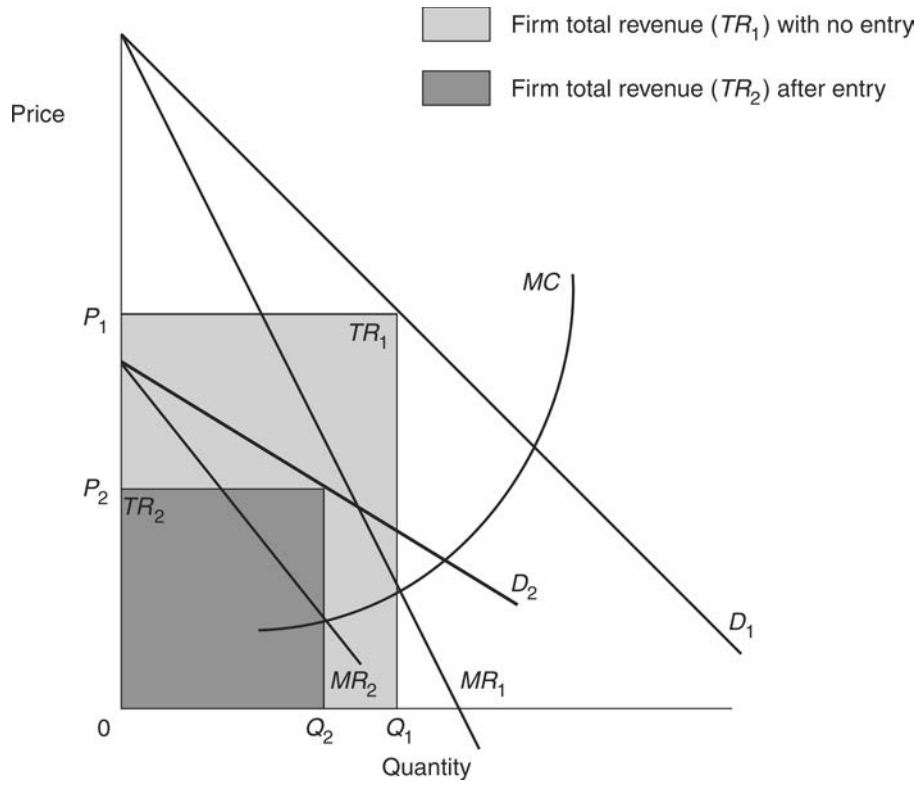


Figure 12.4 Impact of Entry of Alternative Providers into a Monopolistic Market

for each firm (average cost is omitted for clarity). The entry of other firms will have the following effects on each individual firm:

- Shift the demand curve to the left to D_2.
- Shift the marginal revenue curve to the left to MR_2.
- Increase the elasticity of demand at any price because there are now more competitors.

With the same costs facing each firm, the new profit-maximizing price (where marginal revenue MR_2 equals marginal cost) for each falls to P_2. If the decrease in firm demand is sufficiently large, it is possible that an individual firm will no longer be able to earn a competitive return at Q_2. This would occur if the demand curve shifts (due to the entry of competitors) so that it is everywhere below the firm's average cost curve.

The existing firm also may respond in other ways. It may attempt to reduce its administrative costs. More importantly, it may try to court customers by attempting to market plans that limit utilization of services, and hence the costs of the services, through various devices. These include utilization review and the adoption of health care plans with increased cost sharing. Of course, it could establish its own HMOs and PPOs, further increasing competition. It also could improve the quality of its service. Further, it may embrace forms of emerging nonprice competition, such as advertising and marketing.

Is this theoretical scenario applicable? Many items are likely to be affected by the spread of HMOs and PPOs. These include the provider and health insurance markets, the phenomenon of biased selection, the roles of employers, and the rate of innovation and diffusion of technology. To focus our discussion, we will limit it to three areas: (1) the impact of HMOs/PPOs on hospital markets, (2) their impact on insurance markets, and (3) their impact on the adoption of technological change.

Managed Care Competition in Hospital Markets

A first step in looking at the impacts of managed care penetration is to examine the determinants of the penetration itself. What is it about particular markets that lend themselves to high levels of HMO activity? Dranove, Simon, and White (1998) use a demand–supply framework to address this question.

What are the demand characteristics? The authors view employers as the primary health insurance shoppers. Serving as agents for their employees, they seek the best deals.

What are the supply characteristics? The authors seek to identify market features that would reduce costs of payers to contract with providers. They expect that the supply of managed care will be higher in markets where the MCOs can negotiate with and monitor the providers more cheaply. They also expect that excess hospital capacity may make hospitals more amenable to negotiation.

Regarding demand, higher managed care penetration accompanies more educated and more urbanized populations. Regarding supply, higher managed care penetration is related to lower percentages of physicians in solo practice, hospital market concentrations, and hospital occupancy. In short, managed care comes in where large numbers of half-filled hospitals are found. The authors express concern that the low rate of managed care penetration in more concentrated markets may imply anticompetitive behaviors, meriting antitrust considerations.

Early work viewed the hospital as competing for patients, physicians, or both, arguing that widespread health insurance allowed individual consumers and their physician-agents to be much less concerned about the price of care. Hospitals would compete on services and amenities rather than price. McLaughlin (1988, p. 207) argues that the "providers are responding not with classical cost-containing price competition but, instead, with cost-increasing rivalry, characterized by increased expenditures to promote actual or perceived product differentiation."

Feldman and colleagues (1990) examined the Minneapolis/St. Paul area, addressing the degree of competitive bidding for HMO contracts, whether HMOs can get discounts, and whether HMOs tend to rely on low-price hospitals. They found that HMOs generally did not extract major discounts. In fact, price did not seem to be the major HMO consideration in the selection of hospitals with whom to affiliate. Rather, it was hospital location and the range of services that the hospital offered. There was no indication as to whether HMOs tend to affiliate with lower-priced hospitals.

Melnick and colleagues (1992) obtained hospital transaction prices negotiated by a large California PPO in 1987. Their regression analysis controlled for hospital characteristics such as ownership, teaching, Medicare and Medicaid demand, and market structure. Four key findings emerged from the analysis:

1 Controlling for other factors, the PPO paid a higher price to hospitals located in less competitive markets.
2 If the PPO had a larger share of the hospital's business, it was able to negotiate a lower price.

3 The more dependent the PPO was on a hospital, the higher price the PPO paid.
4 Hospitals with high occupancy located in markets with high average occupancy charged the PPO higher prices.

Morrisey (2001) terms the Melnick research "a watershed," demonstrating that traditional concepts of competition could apply to hospital markets and potentially to other health care markets as well.

Managed Care Competition in Insurance Markets

Commensurate with the spread of HMOs, PPOs, and various forms of managed care, the health insurance industry is changing rapidly. Many have noted that health insurers were previously lethargic in introducing innovative insurance products and in their cost-containment efforts. In addition, we consider the historic opposition by organized medicine, tax advantages, and provider control of the Blues, and the constraints imposed by state regulation and antitrust laws, all of which previously inhibited change in the insurance market.

Frech and Ginsburg (1988) identified the dramatic changes that occurred after 1977 when the insurance market was divided about equally between the Blues and commercial insurers. The growth of HMOs and PPOs was accompanied by substantial increases in patient cost sharing, increased utilization review, and self-insurance (or self-funding as described earlier in this chapter) by many large firms.

In a self-funded plan, a Blue Cross and Blue Shield or another organization will act only as a third party in processing claims and providing other administrative services such as utilization review. More competition is introduced because self-insured firms have more control over their health plans and more direct interest in cost-containment measures.

Baker and Corts (1995, 1996) identify two conflicting effects of increased HMO activity on conventional insurance premiums:

1 ***Market discipline.*** HMO competition may limit insurers' ability to exercise market power, thus driving down prices, a standard competitive argument.
2 ***Market segmentation.*** HMOs may skim the healthiest patients from the pool, thus driving insurers' costs and prices up.

Their model suggests that if increased HMO penetration does lower the premium levels, the market discipline effect becomes relatively less important, so at higher levels of penetration, the market segmentation effect may raise premiums. Their empirical work verifies this hypothesis: Across metropolitan areas, market penetration up to 14 percent decreases premiums, but market penetration beyond that point raises them. Wickizer and Feldstein (1995) find comparable market discipline effects, although they do not investigate market segmentation effects.

Joesch, Wickizer, and Feldstein (1998) investigated nonprice impacts of HMO market competition. They found that increased HMO penetration reduced insurers' likelihood of increasing insurance deductibles, or "stop-loss" levels (the levels limiting losses to those insured). Moreover, groups located in markets with higher HMO enrollments were more likely to adopt utilization management or PPO options.

Managed Care and Technological Change

The impact of managed care on technological change is also potentially significant in controlling health care costs. Increased incomes and changed coinsurance rates are insufficient

to explain the increases in health care costs over the past 50 years. Most analysts attribute a major role to the advances in high-cost technologies fed by payment mechanisms that were at best indifferent to controlling costs. To the extent that managed care plans explicitly seek cost containment, one would expect careful monitoring of attention directed toward high-cost technologies.

Baker and Spetz (2000) compiled an index using 18 technologies available in 1983, including cardiac catheterization and neonatal intensive care units. They then aggregated hospitals within metropolitan areas and compared metropolitan areas on the basis of degree of HMO market shares. Fundamentally, they found that HMO market shares did not matter. Although they detected modest variations, no substantive differences were seen in technology at given points in time or in the dispersion of technologies over time.

Managed care plans, through their emphases on cost containment, would seem to be important vehicles for reining in the usage of high-cost-high-technology procedures and facilities, but research has not (yet) proved this. Medical facilities and medical practice styles change slowly, and it may be too early to see changes induced by the recent market penetration of managed care plans. Or it may be that the public demands high (and costly) technology, irrespective of who provides it.

The Managed Care Backlash

In the first half of the 1990s, many managed care plans placed increasingly severe restrictions on patient choices, including prior approval for access to specialists and certain high-cost procedures. The results of a 1997 survey conducted by Blendon and colleagues (1998) documented the public's anxiety about the direction of managed care at that time. Only 34 percent of American adults who were surveyed thought that MCOs were doing a "good job," 51 percent believed that MCOs had decreased the quality of care, and 52 percent favored government regulation even if it would raise costs.

Concerned about timely access to care, California voters passed a law in 2002 intended to ensure that HMO members do not face undue delays in receiving medical attention. The rules were not approved until 2010 by the California Department of Managed Care but their specificity is unprecedented. HMO members will face maximum waiting periods for non-emergency care, e.g., 48 hours for urgent care with no prior authorization, and 15 business days for nonurgent specialty care. While it is too early to assess the effects of the California measures on cost, quality, and compliance, the "drive-through delivery" provides an example that has received extensive media, legislative, and scholarly attention.

"Drive-through delivery," which refers to managed care's movement in the early 1990s toward one-night hospital stays for mothers expecting a normal (or non-caesarean delivery) childbirth, became fodder for late-night talk-show jokes and ridicule of HMOs. Fueled by stories of instances in which children died shortly after the mother's release, there was an enormous public outcry to require managed care plans to provide at least a second night of hospital care. The economics behind this example, however, are serious, and Jensen and Goodman (1999) provide us with an overview.

As recently as 1980, nearly 70 percent of mothers experiencing vaginal delivery had hospital stays of three days or more (Gillum, Graves, and Wood, 1998, Table R). Inpatient care is very costly. Yet, as long as hospitals received reimbursement for what they charged, they had no incentive to send the new mother home earlier. Almost certainly, the marginal benefits to the woman of being in the hospital for a third day did not measure up to the costs of keeping her there.

This cost inefficiency became apparent to HMO managers. Pressure to reduce the stay was considerable, and by 1995, the average length of stay for a mother with a vaginal delivery was 1.7 days, with 46.8 percent of all mothers staying one day or less. For the vast majority, home care (starting the second day) along with appropriate outpatient follow-up has become the alternative. Home care carries its own costs, requiring the woman's time and usually assistance from family, friends, or hired caregivers. Nonetheless, it generally provides an appropriate level of care at a far lower cost than the inpatient care that it replaces.

The key phrase is "normal" childbirth. Medical care is not exact and mistakes are made. Managed care opponents have seized on cases in which a baby sent home the second day after birth developed an ailment and needed to return to the hospital or, worse still, died. Keeping the baby an extra day, they argued, would prevent these problems.

Many states passed legislation requiring insurers to cover at least two nights of hospital stay to all mothers with normal deliveries. Maryland's 1996 Early Discharge of Mothers and Babies Bill guaranteed that mothers and babies have coverage in the hospital for 48 hours for a normal vaginal delivery and 96 hours for a normal caesarean delivery.

Liu, Dow, and Norton (2004) analyze the state length of stay mandates in 32 states, comparing the costs of the increased length of stay to the estimated health benefit, relying on infant mortality estimates that one infant life could be saved for each 1,400 normal newborns moved from early discharge (less than 30 hours) to longer lengths of stay. They find that for normal vaginal deliveries the average state law decreased early discharge (less than two-night stays) by 16 percentage points and increased average hospital charges, implying a $1,281 cost per early discharge averted by the law. This could be converted to a crude estimate of $1.79 million per life saved. They characterize their estimate as neither "highly cost-effective [n]or hugely cost-ineffective relative to estimates of the value of a life (often in the range of US$ 1–10 million)."

Economists have also been interested in another issue closely related to the backlash. Did consumers respond to the restrictions by "voting with their feet"? Marquis and colleagues (2004/2005) examined HMO market penetration in two periods, 1994–1998 and 1998–2001, with the former representing the pre-backlash period. There was little evidence of substantial consumer exit and plan switching even in markets where consumers had more options.

Cooper et al. (2006) provide a more extensive set of results through their analysis of enrollments in HMOs and other types of plans over the period 1997–2003. They analyze both aggregate enrollments as well as enrollments by firm size. Aggregate enrollments remained stable until 2002, well after the most intense backlash years, because enrollment declines in large firms were offset by increases in smaller firms. Beginning in 2002, HMOs experienced sustained aggregate decreases in enrollment.

For various reasons including the moderation in the growth rate of health care costs in the late 1990s, many MCOs eased some of their restrictions. Rather than trying to curtail patient choices and utilization, MCOs have engaged in other strategies to deal with quality and cost containment. One of the most visible is *pay-for-performance* (P4P). P4P describes incentive plans in which physicians and sometimes hospitals are given cash bonuses to meet guidelines for quality care and to implement evidence-based medicine. As a largely untested strategy, the value of P4P is still under debate. Box 12.2 provides evidence of its potential impacts.

We close this section with a revealing report by Fang and Rizzo (2010). There are widespread perceptions that managed care plans have become less restrictive but very little empirical evidence to support this proposition. Fang and Rizzo take advantage of surveys for 2000–2001 and 2004–2005 of large, nationally representative samples of physicians. The proportion of a physician's practice revenues obtained from MCOs was used to measure managed care. The purpose of the research was to compare any changes over the two time periods of managed care and non-managed care in constraining the provision of physician

care. Although the effects of managed care in limiting services in 2004–2005 declined somewhat relative to the earlier period,[9] the bigger story is that it declined relative to non-managed care. The authors conclude (p. 100) that "non-managed care plans appear to be getting more restrictive, similar to their managed care counterparts."

These conclusions should not be surprising. We have already seen that, with the exception of Medicare, managed care dominates other health insurance sectors. It is likely that the remaining non-managed care plans feel the same pressures faced by MCOs, as well as competition from MCOs. Thus most traditional FFS plans have incorporated some managed care features such as prior authorization, general utilization review, or specialty utilization review for mental health and other specialized services.

BOX 12.2

Pay-for-Performance

Rosenthal et al. (2006) documented the extent to which HMOs adopted P4P by 2005. Their survey of a large number of HMOs indicates that more than half (52 percent), representing 81 percent of enrollees, used P4P in their provider contracts, although much more so with physicians than with hospitals. Adoption of P4P was positively associated with HMOs that are nonprofits, those that use primary care physicians as gatekeepers, and those that use capitation to pay them.

P4P is still in an early stage of development, and there are wide variations in program design (Trude et al., 2006). Preliminary results from one of the largest efforts, known as the *Rewarding Results* program, indicated that financial incentives can motivate change if the rewards are substantial. Other evaluations are less encouraging. Rosenthal and colleagues (2005) compared a broad set of quality performance measures for a large health plan that introduced P4P for its California medical groups in 2003, with its plans in Oregon and Washington that did not subject medical groups to P4P. The results for three measures of clinical quality (cervical cancer screening, mammography, and hemoglobin A_{1c} testing for diabetes) showed that the rates for each increased in California after P4P was introduced. However, only the cervical cancer screening rate increased faster than the rate for the Pacific Northwest medical groups.

In a more extensive follow-up to this research, Mullen, Frank, and Rosenthal (2010, p.85) concluded that P4P "may not necessarily have the dramatic or even predictable effects touted by its enthusiasts." For example, appropriate medication rates for asthma even declined after P4P was introduced in California. The effects of P4P on health outcomes were also mixed.

Nevertheless, many analysts remain optimistic about the potential for P4P. The Centers for Medicare and Medicaid Services (CMS) have funded several multiyear P4P demonstrations, and some have shown cost savings and improved patient outcomes. A common criticism of U.S. health care is the lack of financial incentives for quality care. Although this criticism can be debated, the Tax Relief and Healthcare Act of 2006 mandated a P4P program for Medicare. The program, known as the Physician Quality Reporting System, is still voluntary, but the CMS paid an average bonus of nearly $19,000 per participating professional practice in 2009. CMS reported an average increase of 10.6 percent over 2008 across 99 measures of performance.

Accountable Care Organizations (ACOs)

As enthusiasm for HMOs has waned in recent years, policymakers and legislators are rallying around a "new" integrated entity to potentially improve quality and bend the cost curve. These entities are called Accountable Care Organizations or ACOs.

The ACO movement is strongly identified with Elliot Fisher, a professor of medicine at Dartmouth University, who coined the ACO label and articulated its primary features (Fisher et al., 2006; 2012). ACOs have two important elements. First, networks of providers assume collective responsibility for the full continuum of care delivered to a defined population. The providers can be hospitals, community centers, solo practitioners, physician groups, and other entities. The ACO should be led by providers who place emphasis on primary care. Second, the providers are held accountable for quality improvements and cost. They assume risk by sharing in savings (or paying penalties) relative to benchmarks established for cost and patient outcomes. Ideally, sophisticated performance metrics are applied to verify that quality is improving and that any savings are not the result of enrolling lower risk populations.

ACOs, HMOs, and other integrated delivery systems share many characteristics. These include coordination of care across the care continuum, elimination of financial incentives for overtreatment, and an emphasis on primary care and preventive medicine. However, they differ in some important ways. Burns and Pauly (2012) describe some of these differences. ACOs rely heavily on health information technology to support clinical decision making, quality measurement, and disease management; they are subject to performance risk as opposed to insurance risk; and they often use alternative payment arrangements, such as bundled payments for episodes of care, rather than capitated contracts.

Another major difference is size. It is generally difficult for an HMO or other insurance organization to survive unless it can achieve economies of scale and minimize insurance risk. This usually means having tens of thousands of enrollees and, in many cases, much larger numbers. In contrast, ACOs are encouraged to be flexible and innovative. In one study, less than one-half of ACOs were even affiliated with a hospital (Epstein et al., 2014). Together with better coordination of care and tighter scheduling, ACOs have the potential to achieve scale economies with fewer beneficiaries. The Medicare Shared Savings Program established under the ACA requires just 5,000 enrollees.

ACOs are designed to address the shortcomings of both FFS and traditional managed care. We have seen that FFS provides strong incentives for overutilization while HMOs and other managed care entities raise concerns about underutilization. ACOs have incentives to improve quality as well as to reduce costs. Some view them as the epicenter of the drive to value-based care.

Integrated delivery systems with ACO characteristics have served privately insured groups for some time. The current impetus for ACOs is attributable to the Centers for Medicare & Medicare Services (CMS) which projected significant savings if they were allowed to serve the Medicare population. Independent analyses are less clear. Colla et al. (2012) found relatively modest effects in which the substantial savings of some groups were largely offset by higher costs in others. A more general review of integrated delivery systems by Hwang et al. (2013) indicates that the evidence for cost savings "is rather weak" and that, although most studies showed positive quality effects, it is difficult to identify the source of the quality improvements or whether they are even statistically significant.

Despite the lack of compelling evidence favoring ACO performance, the Medicare Shared Savings Program and Pioneer ACOs (a program that places higher risk on providers) were

established under the ACA. The number of organizations in these programs has grown rapidly. About 6 million Medicare beneficiaries belonged to an ACO in 2015. They are among the nearly 24 million that are served by one.[10]

ACOs promise tighter coordination of care and joint financial incentives to providers but Burns and Pauly caution that "care coordination has remained an elusive goal" and that previous hospital–physician partnerships "have not promoted cooperation, improved quality, contained costs, or integrated clinical care" (p. 2410). We conclude by suggesting that ACOs can play a role in health system reform but their potential remains largely untested.[11]

Managed Care and the Affordable Care Act

Coverage of millions of previously uninsured under the ACA is likely to have substantial spillover effects throughout the health care system. Will waiting times increase? Plans must cover ten essential categories of services such as ambulatory care, chronic disease management, and mental health and substance abuse disorders. Will this essential health benefits feature of the ACA strain resources in these categories? The actions and effects of the Independent Payment Advisory Board (IPAB), which will make recommendations for payment reform, are also largely an unknown at this time. We take up some of these issues elsewhere. Here we focus specifically on several ACA provisions that will have immediate and potentially large impacts on managed care providers and their patients.

Employer-sponsored plans under ERISA are minimally affected by the ACA but the legislation has major effects on Medicaid and Medicare managed care. As we have seen, many states contract with private managed care plans for their Medicaid enrollees. With substantially more people becoming Medicaid eligible under the ACA, existing plans will enjoy a boom in demand and new plans might be tempted to enter the Medicaid markets.

States expanding their Medicaid programs must include the ACA's essential health benefits coverage for new enrollees. Some states, such as Michigan, have expanded their program with novel reforms (Ayanian, 2013). Michigan's 2014 expansion, known as Healthy Michigan, covers adults with incomes up to 133 percent of the federal poverty level. It imposes various copayments on most participants with 5 percent cost sharing for new enrollees who fall between 100 and 133 percent of the poverty level. The cost sharing is reduced for those who engage in healthy behaviors and all new enrollees will be placed in private managed care plans with health savings accounts. Finally, if the additional costs of expansion are not offset by reductions in certain related health costs in the state's budget, Michigan will withdraw from the Medicaid expansion as early as 2017.

The Medicare Advantage plans described earlier are also greatly impacted. We have noted that 31 percent of the Medicare population was enrolled in private Medicare Advantage plans in 2015. Most are enrolled in managed care plans that are reimbursed by CMS on a capitated (i.e., per enrollee) basis.

President George W. Bush and many political conservatives favored the expansion of Medicare Advantage as a way of increasing competition and ultimately holding down the growth of Medicare costs. Government payments to Medicare Advantage plans exceeded the cost per regular Medicare enrollee. Political liberals viewed this as a threat to traditional Medicare and, under the ACA, made cuts in payments to Medicare Advantage plans beginning in 2011.

The methodology is complex but reduced reimbursement rates were to be until 2017. At that time, the rates will fall to 95 percent of the costs of traditional Medicare in counties with high costs to 115 percent of traditional Medicare costs in low-cost counties.

In an attempt to improve quality, the ACA also ties payments to quality. CMS developed a five-star system that rates Medicare Advantage plans along five domains (such as managing chronic conditions). Beginning in 2010, CMS made bonus payments to plans that attained at least four stars. Starting in 2015, CMS can terminate plans that have received three stars or less for three consecutive years.

Scholars have yet to evaluate the many changes introduced in both public and private plans. Clearly, change is taking place at an unprecedented rate and CMS is continuing to test other health care delivery models. It is also clear that the rate increase in health care spending is slowing down. How much, if any, of the slowdown is due to the ACA and other recent reforms will be taken up in Chapter 22.

Conclusions

This chapter has considered HMOs and other managed care delivery systems that combine the functions of insuring patients and providing their care. We begin by describing managed care and providing the rationale for a government policy that promotes it.

Our discussion has emphasized that HMOs and other integrated delivery systems have incentives to curtail costs because they serve as both insurers and providers. Thus the incentives for additional and less essential procedures are reduced. We can show that systemwide even less care may be given than would be economically efficient. One key finding is that managed care organizations tend to reduce hospitalization—one of the most expensive components of health care costs. While other findings are mixed, little evidence suggests that the quality of the care provided in HMOs is inferior to FFS care. Another key finding is that MCOs have been able to reduce fees paid to providers.

In the early 1990s, cost pressures moved more users into managed care. By 2015, traditional fee-for-service (FFS) health care enrollment for covered workers had fallen to 1 percent, from 73 percent nearly three decades earlier. Yet customers also rebelled against the more stringent cost controls of HMO plans, preferring what some analysts refer to as "managed care light"—as exemplified by PPO or POS plans. Less stringent controls were palatable with the deceleration of health care costs in the late 1990s. However, costs were increasing again at double-digit rates in the first years of the twenty-first century. The growth of costs slowed down with the onset of the Great Recession but the future remains unclear

The passage of the ACA in 2010 is also creating uncertainties as well as opportunities in both private and public health insurance markets. In addition to the challenge of containing costs, there is growing awareness of widespread deficiencies and inconsistencies in health care quality. These concerns are likely to bring continuous change. The rise of ACOs and mechanisms that reward value-based care are among the recent innovations that could reshape health care delivery.

Summary

1 Managed care seeks to integrate what previously had been a nonintegrated system of health care treatment. Such integration has the potential to reduce health care costs, but the integration is costly and may limit choice of provider and treatment options.
2 In HMOs, PPOs, and other MCOs, the functions of insurance and the provision of care are combined. In return for a prepaid premium, MCOs agree to provide enrollees with comprehensive health care over a given period.

3 By agreeing to handle all of a patient's health care needs for a fixed, prearranged fee, a provider is bearing a substantial part of the financial risk. By bearing such a risk, the managed care organization has a strong incentive to develop strategies for reducing excessive care and minimizing other inefficiencies.

4 HMOs control utilization and costs by imposing physician gatekeepers and requiring treatment within defined provider networks. PPO plans maintain the provider networks but do not require physician gatekeepers. Point-of-service (POS) plans maintain the physician gatekeeper role but do not require treatment within defined provider networks.

5 There has been a dramatic shift from FFS to managed care in employer-sponsored plans. By 2015, only 1 percent of these workers had traditional FFS insurance, 76 percent had some form of managed care, and 24 percent had a high-deductible health plan.

6 A theoretical model of the HMO shows that, due to the impact of potential disenrollment that does not face providers in the fee-for-service sector, one might expect inefficiently low levels of care in the managed care sector.

7 If providers can charge different consumers different amounts, they can earn additional profits. Such price discrimination by providers is difficult under the contracts characterizing prepayment-based organizations because:
   - Providers will find it difficult to determine how much individual consumers value the services.
   - Prepayment-based organizations may be able to shop among providers, thus limiting the providers' monopoly power.

8 Managed care organizations typically provide comprehensive ambulatory and inpatient care, including routine office visits and preventive care, generally with low coinsurance or deductibles. Paperwork for patients is reduced, and uncertainty over their coverage is minimal. These features make membership attractive to consumers, especially to those who are concerned about out-of-pocket costs.

9 In contrast to FFS arrangements, the managed care enrollee's choices of providers and access to hospitals (aside from emergency care) are limited. Also, direct access to specialists may require referral from the patient's gatekeeper—the primary care physician.

10 There is a strong consensus that managed care reduces utilization, especially of hospital care. There is little evidence that the quality of care is inferior to the quality found in an FFS system.

11 The lower expenditures per enrollee under managed care are associated largely with the lower fees they are able to negotiate with some providers. Risk selection is not a major factor.

12 In addition to individual cost-related impacts of managed care organizations, there are also competitive impacts. Although higher market penetration of managed care does not always lead to lower hospital prices, it does appear to reduce insurance premiums.

13 MCOs, through their emphasis on cost containment, would seem to provide important opportunities for reining in the usage of high-cost-high-technology procedures and facilities. Research findings to date have not supported this hypothesis.

14 MCOs adopted more stringent restrictions on utilization in the early 1990s. There was a strong media backlash. Some evidence indicates that consumers did not react by "voting with their feet."

15 Pay-for-performance (P4P) describes incentive programs in which providers, most often physicians, are given cash bonuses to meet quality performance targets. Many MCOs have adopted P4P but the preliminary evidence of its effects is not yet clear.

16 Accountable Care Organizations (ACOs) have emerged as a distinct entity that faces incentives to both improve quality of care and restrain costs.
17 The Affordable Care Act has prompted the rapid growth of Medicare ACOs and brought many other changes to both Medicare and Medicaid.

Discussion Questions

1 What are the key elements that distinguish managed care from FFS plans?
2 What are the principal differences among HMO, PPO, and POS plans?
3 How do the economic profits (rents) that may be earned by some groups of providers enable MCOs to limit expenditures? What role does the price elasticity of demand play in this process?
4 Why is selection bias such an important issue in measuring HMO performance?
5 Discuss ways that managed care organizations may be able to reduce costs of care to their clientele.
6 Why do some critics argue that managed care organizations provide lower-quality care than FFS plans? Evaluate this possibility from a societal perspective.
7 After a large increase in membership, HMO enrollments flattened in the late 1980s and many HMOs suffered financial difficulties. How could this be explained according to what is known about the supply and demand for HMOs?
8 If everyone chose to join an HMO, would average HMO expenditures per case tend to rise or fall? Would national health expenditures tend to rise or fall?
9 What features of managed care organizations tend to inhibit or discourage people from joining? What features tend to attract people? Discuss the advantages and disadvantages of managed care enrollment.
10 Why is the growth of managed care a relatively recent phenomenon? Describe governmental policies and practices that have encouraged managed care organizations and inhibited them.
11 If traditional FFS leads to demand inducement, what constrains the HMO from underproviding care?
12 Explain how the availability of alternative delivery systems is expected to produce competitive effects throughout the health economy.
13 Discuss the ways that managed care organizations can influence the adoption of new technologies.
14 Some critics argue that providers do not have sufficient financial incentives to provide quality care. Describe some of the existing safeguards. Evaluate the potential role of P4P. How are ACOs incentivized to improve quality?

Exercises

1 Consider an HMO with a demand curve of the following form: $Q = 100 - 2P$. Suppose that its marginal and average costs were \$20. If the firm maximizes profits, determine its price, output, and profits.
2 In Exercise 1, if the firm must act as a perfect competitor, in the long run what will happen to equilibrium price and equilibrium output?

3 Consistent with Figure 12.1, assume that the FFS price was $100 per visit and the average patient made eight visits per year. A competing managed care organization came in and charged $80 per visit, providing seven visits per year.
   (a) Calculate the change in total expenditures.
   (b) Graph the FFS and the managed care market equilibria as was done in Figure 12.1. What do our findings suggest about demand for managed care compared to demand for FFS care?
4 Consider the discussion on adverse selection into HMOs and FFS care, as noted through equation (12.2) and Figure 12.3. Suppose that, on average, FFS clients bought $2,000 in services ($\overline{s}_F = 2{,}000$) and HMO clients bought $1,500 in services ($\overline{s}_H = 1{,}500$), with an efficiency factor of 0.9. The FFS plan charges a 10 percent coinsurance rate.
   (a) Set up this problem graphically, labeling the *E* and *V* curves.
   (b) If a client expects to spend $250 on care, will he or she choose an HMO or an FFS plan? Why?
   (c) At which value of *s* would the client expect to be indifferent between an HMO and an FFS plan? Why?
   (d) How would your answer to parts (b) and (c) change if the HMO adopted a 20 percent coinsurance rate?
5 Assume that in Figure 12.4, so many providers entered the health care market that individual demand curves fell below the average cost curves. Draw the new equilibrium. What would happen to short-run profits in the health care market?
6 Exercise 5 discusses a short-run equilibrium in the health care market. With entry and exit into and from the market, graph and discuss the long-run equilibrium.
7 In Table 12.4, the market penetration for managed care rises from 40 percent to 50 percent between Periods 1 and 2 and stays at 50 percent. Suppose, instead, that it fell from 40 percent to 30 percent and stayed at 30 percent.
   (a) What would happen to total costs and to rates of cost increase?
   (b) How do your results compare to the discussion regarding Table 12.4? Why?

Notes

1 Available at kff.org. The KFF sources used in this chapter include: *Employer Health Benefits 2015 Annual Survey; Medicare Advantage 2015 Spotlight: Enrollment Market Update*, June 2015; *State Health Facts: Total Medicaid Managed Care Enrollment*; and *Kaiser Slides*.
2 Employer-sponsored HDHPs with health reimbursement arrangements (HRAs) or health savings accounts (HSAs) were part of the 2003 Medicare Modernization Act. Although these plans often include managed care features, they are usually considered as a distinct category. We elaborate more on these plans in Chapter 22.
3 Following legal challenges to the ACA, the U.S. Supreme Court upheld the constitutionality of the individual mandate that required most individuals to have insurance coverage. However, the Court also ruled that the Act's Medicaid expansion was coercive of the states. The decision effectively made Medicaid expansion optional. As of January 2016, 31 states and the District of Columbia expanded their plans with another three states considering expansion.
4 From a provider's perspective, such as a physician or hospital treating both FFS and MCO patients, its pricing decision is very similar to one involving price discrimination. Chapter 17 formally covers price discrimination within the context of pharmaceutical

products. Here, the provider's demand from the MCO market is likely to be far more elastic than its demand from the FFS market because the MCO can contract with other doctors or hospitals if the provider tries to raise rates for the MCO's enrollees. As shown in Chapter 17, price will be lower in the market with a more elastic demand.

5 Economists have yet to develop a distinct generalized theoretical model of managed care, using either competition or game theory. This section describes some simple approaches that provide useful insights. See Olivella and Vera-Hernández (2007) and Brekke et al. (2010) for more sophisticated contributions that attempt to deal with quality and differentiation across managed care plans.

6 The 2015 CPI was about 2.4 times the 1983 value; the medical care component of the CPI was about 4.5 times its 1983 value.

7 Glied (2000) summarizes many articles that examine impacts of managed care on costs and on utilization. Her tables discuss the populations, comparison groups, controls for differences in patient characteristics, and impacts on treatment aspects, such as charges and length of stay.

8 Competition can also affect quality and other dimensions of care. Scanlon and colleagues (2005) found that lower levels of HMO competition, as measured by the Herfindahl-Hirschman Index, actually produce better results on several quality dimensions. Higher HMO penetration was associated with higher quality. The study also found that plans that publicly report their data have higher quality than those that do not.

9 In another report using the same survey, Fang, Liu, and Rizzo (2009) did not find a change over the two periods in the assignment by HMOs of primary care physicians as gatekeepers.

10 Jeremy Gold, "Accountable Care Organizations, Explained," *Kaiser Health News*, September 14, 2015: http://khn.org/news/aco-accountable-care-organization-faq/, accessed January 15, 2016. The public has little awareness of ACOs and, as Gold notes, "you may even be in one and not know it."

11 In a series of 2015 postings on *Health Affairs Blogs* (http://healthaffairs.org/blog/2015), McClellan, Kocot, and White provide considerable new evidence on the Medicare Shared Savings Program. Their results are encouraging. Some ACOs have actually been able to improve quality while reducing costs.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 13

Nonprofit Firms

In this chapter

- An Introduction to Nonprofits
- Why Nonprofits Exist and Why They Are Prevalent in Health Care
- Models of Nonprofit Hospital Behavior
- The Relative Efficiency of Nonprofits Versus For-Profits
- Conclusions

Nonprofit firms account for only 5 percent of GDP, but they make up a significant portion of the health care sector. The 60 percent of community hospitals that are nonprofit provide nearly 70 percent of the beds and treat a similar proportion of the nation's hospital patients. Nonprofit firms also offer 30 percent of nursing home care and half of the inpatient specialty mental health and substance abuse treatment. We also see nonprofits providing education, collecting and providing blood, operating symphony orchestras and opera companies, and facilitating the conservation of land.

Inasmuch as nonprofits are prominent in health care, especially in the important hospital industry, they pose questions of considerable interest. Will the economic behavior of nonprofits differ distinctly from the more common for-profit firm or from government enterprise? A growing body of literature addresses such questions and adds to our understanding of nonprofit firms in general. Studies within health economics contribute to both the general theory of nonprofit firms and the understanding of vital issues of the health economy.

An Introduction to Nonprofits

What is the nonprofit firm? People commonly assume that nonprofits are firms organized to provide charitable goods or community services and that they obtain their revenues largely from donations. While many fit this category, the definition does not serve well conceptually, nor does it make an important economic distinction. On one hand, numerous profit-making firms provide important goods and services to the community, and many provide basic goods, such as food and housing, to the poor. On the other hand, nonprofit firms often serve the well-to-do, and they often compete with for-profits. Many obtain the lion's share of their revenues from the sale of goods and services at prices similar to their for-profit competitors. Health care nonprofits obtain more than 90 percent of their revenues from "sales and receipts."

In economic language, the important distinction of the nonprofit is the nondistribution constraint. This means that no one has a legal claim on the nonprofit's residual, the difference between the revenues and its costs, or what an ordinary firm would call its profits. Because there is no residual claimant, the nonprofit's objectives may differ from profit-making. Two secondary distinctions between nonprofits and for-profits are also consequential. First, nonprofits are exempt from corporate income taxes and often from property and sales taxes. Second, donations to nonprofits receive favorable tax treatment. These distinctions give them an advantage and make nonprofits potentially different from for-profits, but are they?

Why Nonprofits Exist and Why They Are Prevalent in Health Care

In any economy like that of the United States, there are three types of firms: private profit-making, government, and voluntary nonprofit enterprises. Why do each of these types of firms exist? Societies do not create economic institutions at random. They must reflect some basic economic incentives and unmet needs that the firms and the laws establishing them were created to satisfy.

Nonprofits as Providers of Unmet Demands for Public Goods

Burton Weisbrod's analysis (1975) guides the following exposition. In the United States and other Western economies, economists view a perfectly competitive industry under certain circumstances as economically efficient, and empirical data suggest that competition often fosters growth. Under this account, we need government enterprises only in cases where competitive markets fail. As this theory unfolds, we need nonprofits, in turn, when government enterprise also fails. We begin by reviewing the standard explanation of the role of government enterprise, and we then introduce Weisbrod's explanation (Weisbrod, 1988) in which the nonprofit firm satisfies demands for public goods unmet either by private markets or the government.

Under the standard economic explanation, government enterprise might possibly—though not necessarily—have a role in improving market efficiency in cases where competitive markets tend to fail. The most prominent cases of market failure involve externalities and public goods. Two examples will clarify these issues.

MARKET FAILURE: EXTERNALITIES Most consumer goods provide private benefits and little or no externalities. An externality is an uncompensated direct effect of the production or consumption of a good on persons other than the producers or consumers. Consider goods without any externalities. When one enjoys a hamburger at a local restaurant, the pleasure is primarily private, and the benefit goes to the one who consumes. There is no effect on parties external to the market, other than the producer or consumer. However, consider the case where one purchases a vaccination for influenza. This good entails a private benefit: The purchaser will less likely suffer from influenza. In addition, there is an external benefit to others because the purchaser will be less likely to infect others.

Free markets tend to underproduce goods for which there exist significant external benefits. For example, the purchasers of vaccinations will tend to consider primarily the private benefits and will ignore the external benefits to the community. The benefit to society is the sum of the private benefits and the external benefits to the community. Because demand will represent only the private benefits, it will understate society's benefits and give a false or inadequate signal to the market. The market then produces less than the amount that would maximize net social benefits. This is economically inefficient, and we call this situation a case of market failure.

The existence of a large externality raises the possibility of a role for government, but is it only a possibility? If we recognize that markets may fail, we must recognize that governments too may fail to act efficiently.

MARKET FAILURE: PUBLIC GOODS The vaccination is a private good with an external benefit, but it has public good aspects. Consider first the case of a pure public good, a good that is both *nonexcludable* and *nonrival*. Nonexcludable means that people cannot be economically excluded from consuming the good even if they refuse to pay for it. Nonrival means that one person can consume the good without depleting it for others. Our hamburger is an example of a private good in that McDonald's can easily refuse the hamburger to someone who refuses to pay. Likewise, the hamburger is rival because when one person consumes the burger it is then depleted, eaten up.

In contrast, consider a pure public good like national defense. Imagine an antimissile system that puts a defensive umbrella over a country. This defense system would be nonexcludable because those living in the country would benefit whether they paid or not. It would be impossible, let alone costly, to allow the nonpayer alone to be subject to an enemy missile strike. Likewise, the defense system is nonrival because the protection of one more individual does not diminish the defense enjoyed by others.

The government often provides public goods like this. If private enterprise tried to attempt to provide defense, it would find many citizens choosing to be free riders. A free rider is a person who consumes the public good but refuses to pay. Only government has the power to force consumers to pay.

The Public Good–Private Good Aspect of Donations

Now consider charitable donations. A donation to the health of others has the characteristics of both a public good and a private good. The need for both concepts becomes clear in Richard Steinberg's (1986, 1987) work, which argues that donations to public goods motivate the donor as both private and public goods. If you donate toward the health of a poor person, you may get a "warm glow" (an increase in utility from the act of giving; Arrow, 1975; Andreoni, 1990). The warm glow may come from the act of donating or simply from the pleasure in knowing that a suffering person's health improved. However, other charitably minded persons will also have this pleasure, whether or not they have donated. They are free riders who receive an external benefit free. The charity market then provides too little charity to be efficient.

Now consider a case where the government provides a public good. This case explains Weisbrod's rationale for the existence of nonprofits. In Figure 13.1, let the curves D_1 through D_5 represent the demand curves of five different voting individuals for a public good that the government will provide. Let the demand curves represent the external benefits to these different groups of taxpayers. These demand curves represent the marginal benefits to the taxpayer donors.

To pay for this public good, assume that these five different taxpayers will be equally taxed at a per-unit tax rate of *MT*, the marginal tax. Because the five individuals will pay equal taxes, the marginal tax for each will be exactly one-fifth of the marginal cost to society. Thus, if the government were to provide output 0*C*, then each of the taxpayers would be

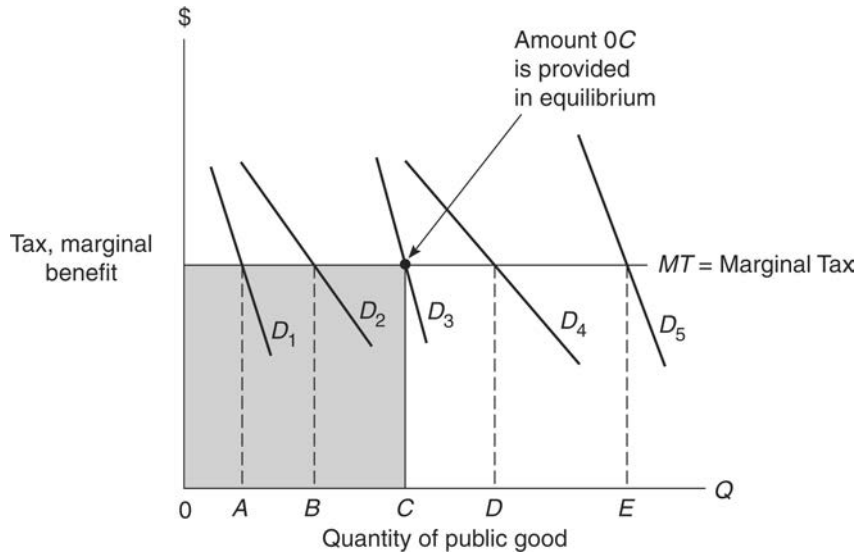


Figure 13.1 The Median Voter Model of Public Goods

charged $0C \times MT$, and the total tax receipts would exactly pay for the project. If the output were $0B$, then $0B \times MT$ would be collected from each individual, and so on.

The government must choose a single level of output. What level would it choose leaving the outcome to the democratic political process? Economists find it convenient to imagine an elected manager whose continued term in office is determined by the ability to provide the "right" level. If level $0A$ is proposed, four of the five voters would prefer more (why?) and would vote against it. If level $0B$ is proposed, three of the five would prefer more and would vote against that level. In contrast, level $0D$ would provide too much public good for three of the five voters, and similarly for level $0E$. Only level $0C$ will gain a majority vote. This majority consists of Voter 3, who is exactly satisfied because her marginal benefits equal the marginal tax, and Voters 4 and 5, who would like to see more but are satisfied with level $0C$. Suppose the manager proposed just a little more than $0C$. This is too much for Voter 3, who now votes against the manager in favor of someone else who would reduce the public good level back to $0C$.

This example illustrates Weisbrod's point. The level chosen entails dissatisfied voters on the margin, whose demands are not exactly satisfied. The marginal tax rate perfectly matches the preferences only of Voter 3, the median person. It is theoretically possible to design a tax system so that for each person the marginal tax equals his or her marginal benefit, but it is difficult to accomplish this in practice. Without such a tax system, some voters will prefer less of the public good and some will prefer more of it. The government enterprise will have failed to satisfy the demands of those who would prefer more. Such unsatisfied voters will have the incentive to form a nonprofit enterprise and provide the good themselves.

It follows from Weisbrod's theory that nonprofits arise because both private markets and government will tend to underproduce goods or services entailing beneficial externalities. The underserved citizens are those who feel the external benefit most keenly. Such will found and support nonprofit corporations.

Relevance to Health Care Markets

The Weisbrod analysis will apply in principle to services that provide external benefits to the community at large. Nonprofit health care enterprises may arise where a sufficient minority of voters are dissatisfied with the quantity or quality of such services provided by the for-profit sector or government. The theory fits the historical rise of nonprofit hospitals. Hospitals in the United States and in many countries often began as charitable institutions, providing care primarily to the poor and relying heavily on donations.[1]

However, the purely charitable aspects of hospital care have become less important (see Gruber and Hungerman, 2007). Improvements in medical technology transformed hospitals into workshops for doctors—places for treating all patients, both rich and poor. By 1996, donations accounted for less than 3 percent of average hospital revenues, and only about 3 percent of patients relied on donations or welfare as their principal source of payment. Hospital finance departments wrote off unpaid patient bills, but these accounted for less than 1 percent of billings. Experts have attributed the demise of hospital donations to the growth of hospital insurance coverage, especially since the advent of Medicare and Medicaid. The decline in donations is consistent with the Weisbrod theory. Individual demands have stayed stable while increases in public sector provision or financing have occurred.

With the decline in cash donations, other forms of donations as well as other forms of nonprofit hospital advantages have become relatively more important. These include donations of time and energy for board members and others.

To be sure, cash donations still play an important role in certain capital fund drives, but in the present day, the financial advantages enjoyed by nonprofit hospitals stem more

importantly from the subsidization of nonprofits by the government through tax exemptions. Generally, nonprofits do not pay corporate, property, or sales taxes, although for-profits do. The only tax advantage enjoyed by for-profits is the ability to write off losses during bad years.

Nonprofits as a Response to Contract Failure

A related theory with early origins complements Weisbrod's ideas. Arrow (1963) suggested that the prevalence of nonprofits is due to the uncertainty of identifying quality of care. Hansmann (1980) expanded this idea and theorized that the nonprofit sector helps to repair the problems of contract failure that occur when the quantity or quality of output is difficult to observe. Thus, asymmetry of information between the firm and the buyer of services becomes important (as noted in Chapter 10) in explaining the nonprofit role.[2]

A classic example of the contract failure illustrates the problem. Suppose you wish to contribute food and clothing to suffering people in Haiti. You can find a firm to deliver the care. However, it would be prohibitively costly to verify that the firm actually is delivering the desired goods to the designated population. You would hardly fly to Haiti just to check on this. Under these circumstances, you may prefer to employ a nonprofit firm.

People will perceive a for-profit firm to have a conflict of interest as a deliverer of the aid packages. Such a firm could increase its profits by reneging on its promise. The nonprofit cannot distribute its residual so it would have less incentive to renege. In this case, the nonprofit, by better serving the donor's interests, also serves the market more efficiently.

Applications of Contract Failure to Health Care

Contract failure does not occur solely in the cases of donated goods and services, as it can exist even when the purchaser is nearby. Contrast the management of hotels and nursing homes. Hotels are profit-seeking enterprises that provide rooms and suites along with housekeeping services, dining, and recreation to travelers and residents. Nursing homes also provide rooms and suites, dining, and recreation along with housekeeping services and of course special nursing services largely for an older and often infirm population. In fact, many nursing homes began as hotels.

We assume that hotel patrons are utility-maximizing consumers who compare benefits and costs in deciding whether to stay at the hotel. Nursing home patients, in contrast, may not be able to assess the quality of the facility and the care they receive accurately, perhaps because of their health impairments. Relatives or friends may obtain only limited impressions upon visiting and may not be sophisticated assessors of the quality of long-term care. A for-profit home, in contrast to a for-profit hotel, may appear to have a conflict of interest in the eyes of some demanders. That nonprofit ownership is viewed as a signal of higher quality is supported by both theory and data, and Christensen and Arnould (2005) provide an example of this. But do for-profit nursing homes in practice provide lower quality? It is difficult to compare them. For example, does a nursing home with more nurses per patient provide higher quality or is it simply a measure of inefficiency? Outcome measures, such as changes in patient health status, are best in principle, but reliable outcome measures are difficult to obtain.

Much of the early evidence suggested an apparent for-profit advantage in cost. However, other reports that took greater effort to account for quality disagreed with the earlier view. In practice, nursing homes with more personnel per patient tend to have better patient outcomes, but it is difficult to sort out efficiency and quality. Gertler (1989) addressed these problems and found a higher quality level in nonprofit homes.[3]

Hirth (1999) and Santerre and Vernon (2007) demonstrated that under plausible conditions, an influx of nonprofit homes will drive up the average quality in the market, making the nonprofit a productive agent for change whether or not it exhibits a higher quality itself. Related research further supports this theory (Grabowski and Hirth, 2003).

Contract failure theory does not appear appropriate for hospitals, a point noted by Hansmann (1980). Hospital patients are under the close supervision of physicians acting as the patients' agents. Physicians have little incentive to misinform patients by overstating the quality of care.

Sloan (1988) further suggests that those who favor the contract failure theory often apply a double standard. The field of physician services, which are nearly all provided on a for-profit basis in the United States, seems ripe for the application of contract theory. Consumers would find it just as hard to assess the quality of physician care as they would hospital care, so why do nonprofit firms not take over the physician care sector?

Financial Matters and the Nonprofit

Lacking the ability to distribute net revenues, the nonprofits do not issue equity stock and lack this avenue for raising capital. Although this is a disadvantage when it needs to respond to changing market conditions, the nonprofit also has some financial advantages. It is exempt from corporate, property, and sales taxes, and its bonds are generally tax-exempt as well. It may also have market advantages, such as the responsiveness and loyalty generated by the consumer's response to nonprofits under the contract failure situation. Finally, it is more likely to attract donations than is the for-profit.

What is the bottom line when all advantages and disadvantages of the nonprofit are considered? Which ownership form can respond to a rapidly expanding demand the quickest? Research generally has indicated that rapid demand change favored the for-profit as measured by changes in their market share. Hansmann, Kessler, and McClellan (2002) determined that over the previous 20 years, the for-profit hospital form was the most responsive to demand change in an era of declining hospital demand. Chakvarty et al. (2005) support this assessment by finding for-profit hospitals to be quicker to either enter or exit a market as conditions change.

Summary of the Reasons for the Prevalence of Nonprofits

Weisbrod accounts for nonprofit firms that arise to provide for unmet demands for public goods when there are significant external benefits from the good or service. Hansmann's view complements this account, emphasizing the role for nonprofit firms in cases of contract failure. Under either of these analyses, the tax preferences for nonprofit firms make economic sense.

Models of Nonprofit Hospital Behavior

Health economics offers many descriptions of hospital behavior. We begin with a nonprofit model that applies in principle to an entire class of nonprofit firms.

The Quality–Quantity Nonprofit Theory

When economists model a nonprofit hospital, they begin by positing an objective of the hospital decision makers. Most frequently, they choose either a utility-maximization or a

profit-maximization type of model. The utility maximizing model, most clearly approximating the altruistic firm, was proposed by Joseph Newhouse (1970). For Newhouse, the hospital's objective was to maximize the utility of the decision makers. Utility of the firm is an index of the decision maker's preferences, a measure of satisfaction, similar to the utility of a consumer.

It is especially difficult to describe the complex elements and interactions of the heterogeneous set of hospital decision makers. Nonprofit hospitals in the United States tend to have three parties with considerable decision-making authority. The *trustees* are nominally in charge, but boards of trustees may include people with widely varying backgrounds, knowledge of health care, and management expertise. The trustees' decision making agent is the *hospital administrator* or CEO. This manager may have varying degrees of actual power and authority. Finally, the arbiters of medical decision making are the *physician staff*. The physicians also may exercise considerable decision-making authority. We assume that this complex decision-making apparatus resolves into a single utility function and describes a set of well-behaved indifference curves.

THE UTILITY FUNCTION In Newhouse's model, the hospital's preferences are defined over quantity and quality of output. Quantity of output could be measured in several ways, but assume we measure it by the number of cases treated. We further assume that there is only one type of case to treat, though there could be hundreds. Output quality can entail many different characteristics of the care provided. Some top decision makers may value the quality or beauty of the hospital structure, and others may emphasize expertise of the physician or nursing staff. Still others may emphasize prestige in the medical community, and yet others may stress the quality of the tender loving care provided. Graphically, we shall suppose that just one index of quality exists.

This conception of the nonprofit hospital is consistent with the external-benefits account of the role of nonprofit firms. The utility derived from producing quantity and quality might arise because care to these patients entails an external benefit to the community at large. Consider this model as a description of the hospital decision makers' having altruistically internalized the community benefit in providing quantity of care.

THE QUANTITY–QUALITY FRONTIER The hospital selects a combination of quantity and quality that maximizes utility. It faces a budget constraint, as it must pay its bills and cannot run negative net revenue. Furthermore, the nondistribution constraint, which applies to all nonprofits, implies that this hospital has no incentive to maximize net revenues as a general rule. Thus, by the budget constraint the sum of patient-generated revenues plus donations equal the hospital's costs.

Figure 13.2 shows its possible choices as the Quantity–Quality Frontier. This frontier comes from demand and cost analysis, but we can understand it intuitively as follows (see Spence, 1973, for a detailed explanation). At zero quality, point Q^*, this hospital can achieve both a higher quality and a higher quantity by choosing a point to the northeast. This can occur if the higher quality attracts more customers but costs do not grow as rapidly. The frontier eventually bends backward, indicating that quality improvements no longer strongly attract customers but quality is increasingly costly.

MAXIMIZATION OF UTILITY Given the possibility frontier, the hospital decision makers choose the point that maximizes utility. The constrained utility maximization point occurs at a point of tangency between the frontier and the highest indifference curve attainable. In Figure 13.2, utility maximization occurs at point *A*.

On the one hand, let the hospital preferences place sole weight on quantity. Then, the hospital would behave like a constrained quantity maximizer and produce at point *B*. On the other hand, let the hospital primarily value quality, this quality emphasis by top decision

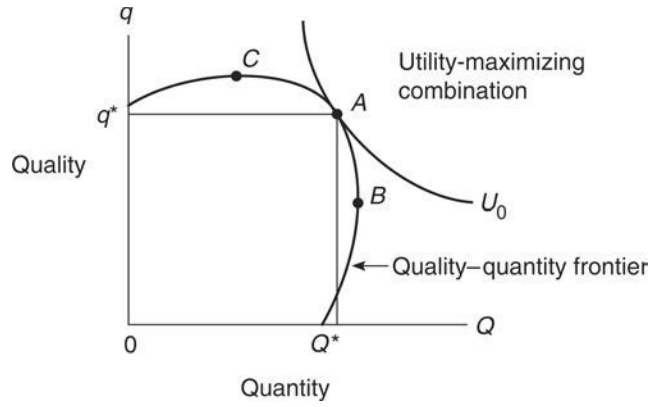


Figure 13.2 Hospital Quality–Quantity Trade-Offs

makers who may be altruists or who may alternatively be primarily interested in the prestige of the hospital as compared to its peers. It would produce at point *C*.

The Profit-Deviating Nonprofit Hospital

We have just seen a model where the nonprofit focuses solely on the quality or quantity of hospital care. Although some have suggested that such behavior might be a form of altruism, Lakdawalla and Philipson (2006) see the nonprofit differently, as a mix of altruism and profit motives. Their model makes clearer the entry and exit responses of nonprofits to changes in market conditions and government regulation.

Let the hospitals maximize utility $U = U(q, \pi)$ over the quantity of hospital service, q, and define the nonprofit's "profit," π^N, as the sum of profit from sales, π_S, and receipts from donations, D, and require that it can pay its bills, $\pi^N = \pi_S + D > 0$. This model is particularly useful to contrast the behavior of nonprofits and for-profits. Thus let the for-profit goal be to maximize its profits π^F.

Notice that the profit-deviating model described this way includes the purely altruistic model (a version where the decision maker has no concerns for personal profit) and the pure profit-maximizing model (where decision makers have no concern for the health of the community separate from profit) as special cases.

Consider entry and exit of the profit-deviating hospital. To enter a market, it must cover its opportunity costs elsewhere, that is, it must be able to attain the utility level that it could achieve in other markets (or by simply not producing in this market). We call its minimum required utility in this market the *Reservation Utility*, $U^* = U(0,0)$. As in all such models, the hospital of both types must pay its bills; here the profit-deviating hospital has an advantage over the pure for-profit because it receives donations. Thus the operating constraint of the for-profit hospital is that its profits are nonnegative, $\pi^F > 0$, while the operating constraint of the profit-deviating hospital is $\pi^N = \pi_S + D > 0$.

Figure 13.3 illustrates their differences. The first graph shows the for-profit's Entry Conditions Curve, which is the same as its Long-Run Average Cost Curve (LRAC). It records all price quantity combinations that yield a zero profit; it must attain at least this much revenue

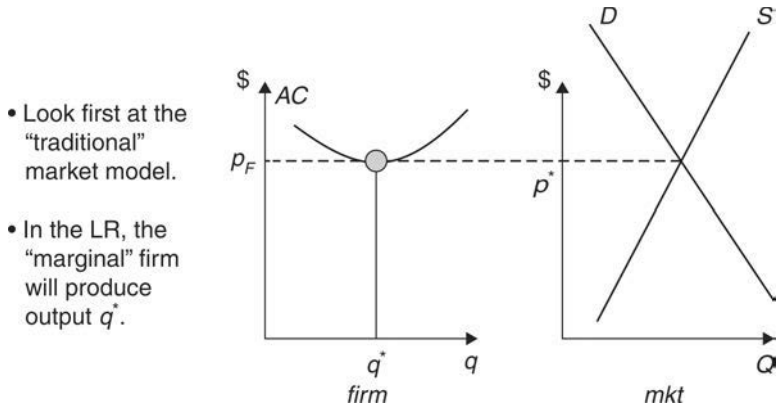


Figure 13.3 Entry Conditions into the Market for Nonprofit and For-Profit Firms

to stay in the market. This much is as in conventional microeconomic theory, and in the long-run equilibrium, the market price would equal P_F.

Alternatively, consider the nonprofit. It too has a break-even curve, though this does not determine its entry conditions. This break-even curve is found by subtracting the average donations from the LRAC curve shown, that is, nonprofit must make at least LRAC – D/q to be able to pay its bills. It will do better. Consider the curve labeled $U(\pi,q)=U^*(0,0)$, in Figure 13.4. This is an indifference curve of a sort, with all points on the curve yielding the same utility. At the utility level shown, the firm is just indifferent between producing in this market or not.

We finish up by explaining the implications of the model. First suppose some more nonprofits besides the one shown enter this market. This would shift the market supply curve to the right and cause the market price to fall. If enough nonprofits entered this could drive the market price below P_F, which would drive all of the for-profits firms out of the market. The nonprofit donations are a very powerful advantage. Why do we have any markets at all with a mix of for-profits and nonprofits?

Keep in mind that these nonprofits will differ in their level of Reservation Utility (much like consumers will differ in their appreciation of a consumption good). This will mean that the number of nonprofits that choose to enter cannot be determined a priori; though, in principle, there could be so many as to drive all for-profits out. Most states in the United States do not have for-profit hospitals, though this must derive in part from legal or other institutional restrictions.

If there are enough nonprofit hospitals waiting in the wings, their entry would drive out all the for-profits. However, suppose profit-deviating entry stops before that so that temporarily there is unmet demand for hospital care. In contrast to the nature of the nonprofit entry, standard for-profit theory suggests that there will always be for-profit firms to take advantage of profit opportunities. However, in order for for-profit hospitals to enter the market, available revenue must be high enough for them to cover their costs. Given unmet demand, the market price will tend to rise until it is profitable for them to enter (in the process some new nonprofits may also enter). The equilibrium is determined by the for-profits' greater response to market signals. In the graph, this price is P_F.

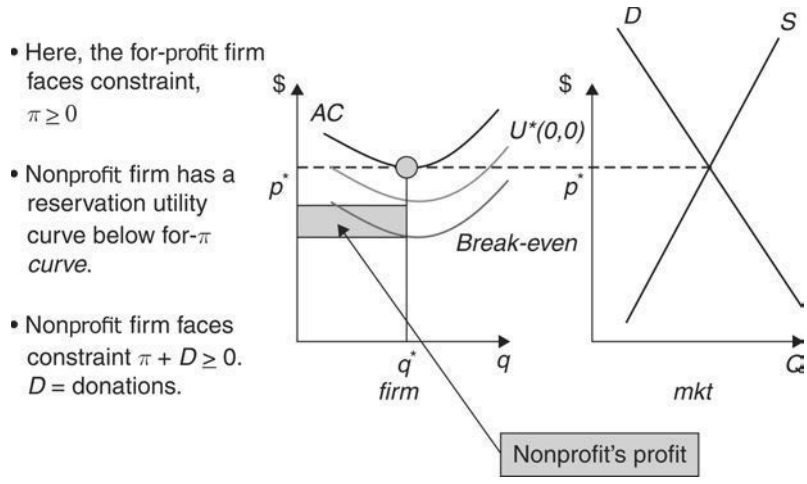


Figure 13.4 Co-Existence of Nonprofit and For-profit Firms

Finally, if prices went up further, then more for-profits would enter. Their entry, as in the standard theory, would drive the price back down to equilibrium. By this process, in the long run, the for-profits are the marginal firms and their cost structure determines the market equilibrium price. Because they are the marginal firms in the industry, they are the ones that determine a new market price after a new regulation changes hospital cost structures. The authors conclude that in markets where the two firm types coexist, the for-profits, as the marginal firms, will determine the effects of the new regulation.

The Hospital as a Physicians' Cooperative

A different account of the nonprofit hospital comes from theories that believe that hospitals maximize the pecuniary gain to the decisive set of decision makers. Mark Pauly and Michael Redisch (1973) describe the nonprofit hospital as a "physicians' cooperative," assuming that the hospital is controlled by a physician staff who operate the hospital to maximize their net incomes.

This view of the hospital focuses on the "full price" of the hospital care, meaning the total charges to the patient by both the hospital and the physician. Assume that the patient pays a single bill representing the full price of the care, and let donations be zero. The full price of care depends on consumer demand and on the total quantity of care offered by the physicians' cooperative. The amount of care produced and offered to patients depends, in turn, on the quantity of inputs chosen by the top decision makers, here the physicians. Summarize hospital inputs as capital, K, and labor, L. Let the physician input be M, a fixed input if the hospital selects a "closed staff."

Maximizing Net Revenue per Physician

In this model the cooperative runs the hospital to maximize the net revenue (NR) per physician (M), or NR/M. The net revenue is the sum of all of the revenue less factor payments to nonphysician labor and payments to capital. The net revenue per physician divides that revenue

over the total number of physicians, *M*. An increase in the number of physicians, *M*, initially increases revenues per physician. Eventually, revenues per physician must fall because (due to the fixed levels of nonphysician labor and capital) the percent increase in revenues (in the numerator) will be smaller than the percent change in number of physicians (in the denominator).

Figure 13.5 shows the optimal staff size if the physicians can limit the size of the staff, or "close" the staff. Here dollar values are on the vertical axis and the number of physicians on staff, *M*, is on the horizontal axis. The curve *N* denotes average physician income. The *N* curve starts at the origin point *A* (no revenue), rises to a maximum at point *B*, and then falls. Curve *s* depicts the supply curve of physicians, which is infinitely elastic, plausibly representing a case in an urban, physician-dense environment. For physicians who are on the staff, the optimal staff size would be M^*, where curve *N* reaches its maximum.

In contrast, if the hospital has an open staff, physicians are free to enter as long as their resulting average income, *N*, equals or exceeds their supply price, *s*. The open-staff equilibrium occurs at point *C*, where net revenue (the demand for physicians) equals supply, s_0, with *M* physicians hired. Regardless of the number of physicians on staff (either a closed-staff or an open-staff equilibrium), the hospital inputs are chosen to maximize residual income for the medical staff. If we view the Newhouse model as resulting from the maximization of external benefit perceived to accrue to the community, then the Pauly-Redisch objective is the complete opposite.

A Comparison of the Quantity–Quality and the Physicians' Cooperative Theories

It is useful to contrast the two models on the extreme ends of the spectrum by comparing them on the same graph. To do so, we represent combined (physician and hospital) revenues as a single function:

$$R = R(K, L, M_0)$$

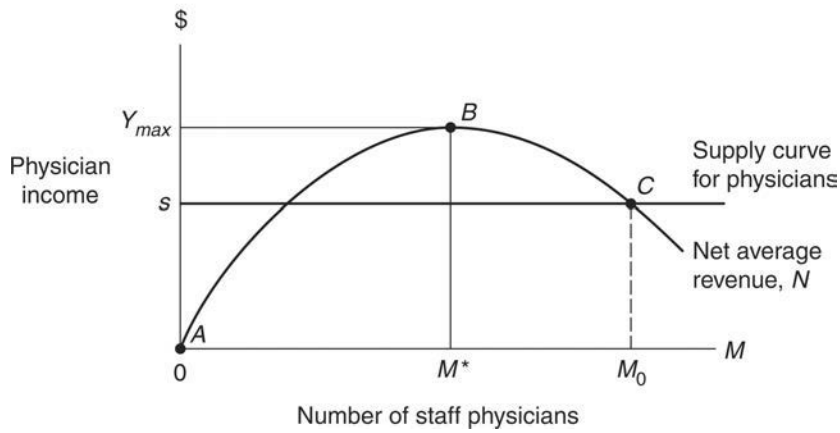


Figure 13.5 Maximizing Net Average Physician Revenue in Hospital

The combined revenues depend on the quantity and quality of care produced, which in turn depend on the amounts of the inputs used. Like Newhouse, assume that the hospitals/physicians produce care efficiently so that a higher quality of care requires necessarily a higher level and hence cost of inputs. The hospital may receive additional revenues in the form of donations, D_0, and government subsidies, G_0. Let physician supply be perfectly elastic at a constant supply price, s. Constant input prices, r for capital and w for labor, complete the description. Finally, define the hospital residual, HR, revenues as the following equation:

$$HR = R(K,L,M_0) - wL - rK - sM_0 + D_0 + G_0 \tag{13.1}$$

Under the Pauly-Redisch model, the physicians on the staff usurp the hospital residual, HR. For a given level of physicians on the staff, M_0, the physicians will maximize their average incomes by maximizing HR itself.

In contrast, the Newhouse hospital will maximize utility of quantity, Q, and quality, q, subject to the constraint that the hospital residual is zero; that is, the hospital breaks even. To simplify the graphical representation, let the hospital residual function, HR, form a rounded hill, following Spence (1973). The contours of that hill in Figure 13.6 graph quantity and quality of care on the axes. The contours are analogous to contour lines on a topographical map. For example, the contour line labeled $HR = 1$ represents the collection of all combinations of quality and quantity of care that yield a hospital residual of \$1 million. Contours farther away from the maximum residual point, $HR = HR_{max}$, yield successively lower levels of hospital residual. The contour curve labeled $HR = 0$ indicates the combinations of quality and quantity that yield a zero residual.

QUANTITY–QUALITY CONTOURS AND HOSPITAL RESIDUAL Begin with the Pauly-Redisch analysis of a closed hospital staff. Let the hospital physician staff be fixed at some level, M_0. The Pauly-Redisch hospital chooses the quantity–quality combination that maximizes the hospital residual, point $HR = HR_{max}$. The hospital residual then goes to the staff physicians, and these physicians will thus have maximized their average incomes. In contrast, the Newhouse hospital maximizes utility at point A, the point of tangency between the $HR = 0$ contour, representing the Newhouse budget constraint, and the highest

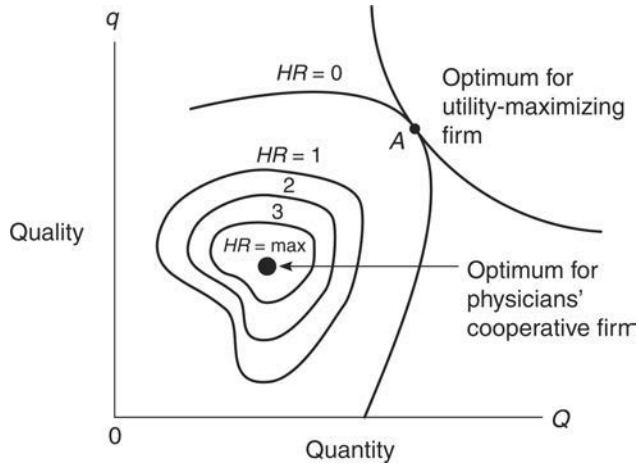


Figure 13.6 Hospital Quantity–Quality Contours and Hospital Residual

indifference curve attainable. Thus, in the closed-staff case, the models yield very different results. As depicted, the Newhouse hospital tends to produce more quantity and quality of care. Under the Pauly-Redisch behavior, however, the physicians indirectly usurp the hospital care residual, and this includes the donations and government subsidy as well. It is as if the nonprofit hospital is a for-profit firm in disguise. If we believed that the nonprofit hospitals behaved like this, we would likely call for an end to government tax exemptions for this nonprofit status (Clark, 1980).

EFFECTS OF INCREASED COMPETITION Examine an effect of increased competition in the hospital sector. If entry were free, then all potential firms that may want to compete for hospital care patients are free to do so. Potential competitors could include alternative delivery systems as well as other hospitals. As more competitors enter the market and compete for business, the demand for care at any existing hospital will tend to fall. For our purposes, this means that competition will tend to shrink the hospital residual hill in size.

Figure 13.7 depicts a case where competition has continued until the maximum hospital residual attainable is zero. As we move away from the top of the hospital residual hill in any direction, the contours reflect negative and successively more negative residuals. The result is that the Newhouse and the Pauly-Redisch nonprofit hospital in this result will converge in their choices of quantity and quality. The only difference remaining in the long run will be simply the result of the nonprofit's cost advantage.

Competition from Home Care and Outpatient Care

The hospital industry has experienced increasing competition in the last 30 years. Especially noteworthy is the competition from alternative delivery systems. For example, home health care can substitute for inpatient care and is frequently cited to be more cost effective. Yoo et al. (2004) have shown that home care and long-term care are close substitutes. In a different sense outpatient care which substitutes for inpatient care can be thought of as a competitor. A study by Vitikainen et al. (2010) employs Finnish data finding that total system costs are

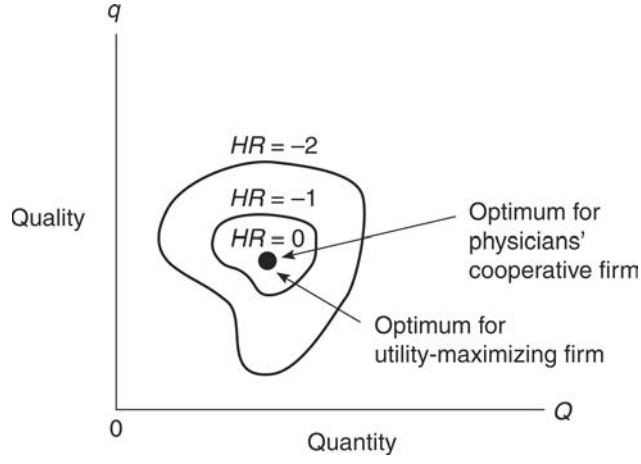


Figure 13.7 Convergence of Hospital Models in Long-Run Equilibrium under Free Entry

not very different under the two regimes. Surprisingly when the authors applied stochastic frontier analysis, outpatient care revealed a reduced efficiency in production.

The Evidence: Do Nonprofit Hospitals Differ from For-Profit Hospitals?

Since the physicians' cooperative would choose nonphysician inputs and outputs much like a pure profit-oriented firm, we look at empirical comparisons of nonprofit and for-profit hospitals. Despite what appears to be a big difference in orientation, only a few behavioral differences are clear in theory. Sloan and colleagues (1998), in fact, conclude that there is not a "dime's worth of difference," basing this conclusion on studies of quality, cost, and efficiency of hospitals by ownership type.

Theoretically the nonprofit will produce at a higher quality, a higher quantity, or both. The Newhouse nonprofit could be very quantity oriented; it may strive to provide large quantities of basic hospital care to a deserving, perhaps destitute, population. Careful observation, however, finds that the public hospitals are first to serve areas of poverty (Ballou, 2008). Nonprofits may instead strive single-mindedly for the highest quality of care possible (Newhouse argued that there would be a bias toward quality). These (potentially) different types of nonprofits make comparisons of average quality of all nonprofits versus for-profits difficult to interpret. Thus, lack of quality differences on average is consistent with theory.

Are there differences empirically in quality? An early study by McClellan and Staiger (1999) found higher mortality rates for the elderly in for-profit hospitals overall, but the small difference on average masked substantial variation with a number of markets showing quality superiority in the for-profit hospitals. Shen (2002) and Lien et al. (2008) also found quality advantages with the nonprofits, though Eggleston et al. (2008b) warn that these kinds of results depend on data sources, time periods, and regions studied. There is mixed evidence on the effect of competition on quality (Gaynor, 2006), but at least one study finds that competition from nonprofits tends to provide spillover effects so as to improve the quality of the for-profits (Grabowski and Hirth, 2003).

Who provides the most charity care? Norton and Staiger (1994) measured differences in care provided to the uninsured by nonprofit and for-profit hospitals. They found that hospitals in the same market area tend to serve the same number of uninsured. Private charity care declined historically, probably due to crowding out by government charity that started with President Franklin Roosevelt's 1933 New Deal (Gruber and Hungerman, 2007). It may be possible, nevertheless, to encourage nonprofit hospital markets to increase their charity care. This could occur if the financial incentives of tax-exempt status encourage development of nonprofit hospitals, which, according to Hassan and colleagues (2000), "are forced to provide higher levels of charity as a condition for gaining access to the tax exempt [bond] market."

Several studies investigate potential differences in managerial practices. Ballou and Weisbrod (2003) find substantial differences among religious, secular nonprofit, and government hospitals in patterns of CEO compensation. However, Brickley and van Horn (2002) find for a large sample of nonprofit hospitals that compensation incentives for CEOs are significantly related to financial performance. They also find little evidence that nonprofit hospitals provide "explicit incentives for their CEOs to focus on altruistic activities."

Several technical issues have shown the differences between for-profit and nonprofit hospitals more clearly. As we discussed, Hansmann et al. (2002) found the for-profit to be quicker in adjusting to market demand changes, presumably due to better access to the capital markets. Chakvarty et al. (2005) support this finding that the for-profits are more "nimble" in adjusting to new economic conditions. Hirth et al. (2000) studied the responses of firms

providing dialysis care during a period of decreasing dialysis payments. The nonprofits tended to maintain quality and offset the dollar losses by cutting back on amenities, while the for-profits took the reverse route.

Zalecki and Esposito (2007) examined possibly different responses to market power. They avoid the more common approach to measuring changes in price given market power. For-profits tend to raise price, but the authors found that price was an inadequate measure. Instead they estimated the change in admissions. They found that nonprofits differ from for-profits and publics by "ignoring market power altogether when determining capacity utilization."

Finally, Silverman and Skinner (2001) found different patterns of DRG "creep" in which hospitals may "upcode" or recategorize patients into more profitable Diagnosis Related Groups (DRGs) used to determine Medicare payments. The authors compared the DRGs of "pneumonia" and "respiratory illness"—where the line between the two DRG categories is fuzzy for many patients. The authors found generally that hospitals facing tougher financial conditions tended to upcode patients into the more profitable DRGs. The upcoding response, however, was greater for the for-profit than for the nonprofit hospitals, and greatest for hospitals that had just recently converted to for-profit status.

Summary of Models of Hospital Behavior

While many theories of nonprofit hospital behavior exist, we have described some main types. Most commonly, researchers depict hospitals as either utility maximizers or profit maximizers. Of the utility-maximizing type, the Newhouse model is a prominent example, and it depicts the top hospital decision makers as choosing the best combination of quantity and quality of care. The middle ground is occupied by the Lakdawalla-Philipson theory, which proposes that nonprofit preferences include altruism and profit maximization. Of the profit-maximizing models, the Pauly-Redisch physicians' cooperative version is most prominent. This approach depicts the nonprofit hospital as effectively making choices that serve to maximize the pecuniary gain to physicians—the decisive set of decision makers.

The continued attempts to distinguish the theoretical models will remain a serious interest for policy in debating the nonprofit's favored status. For example, would we continue to extend tax-exempt status to the Pauly-Redisch type of nonprofit? Ironically, if we become successful in identifying nonprofit hospitals by behavioral type, we may find that hospitals of all types coexist.

What Causes Conversion of Nonprofits into For-Profits?

About 7 percent of nonprofit hospitals converted to for-profit status between 1970 and 1995, and the conversion rate has increased in recent years. Usually, conversion means the sale of the hospital assets to a profit-oriented corporation and the use of the proceeds to fund a nonprofit foundation. Several ideas recur in this new literature. As we have seen, Lakdawalla and Philipson (1998, 2006) explain that if nonprofit hospital decision makers value profit goals and output goals, they will convert to for-profit status when opportunities for making and enjoying profits provide greater utility than any other combination of quantity and profits. Several Blue Cross organizations have converted in this way.

Thus increases in consumer demand that improve profitability will increase the chance that some nonprofits will convert. In such cases, shedding the nonprofit constraint has become more attractive. Others make essentially the same observation in noting that the secular decline in the importance of donations to nonprofit hospitals makes continued nonprofit status less important to many decision makers. A reduction of tax benefits can also erode

loyalty to the nonprofit status; the data show that tax rate changes have the corresponding effect on nonprofit market share.[4]

More idiosyncratic, individual motives may be just as important. Financial distress frequently motivates conversion, and the new management often succeeds in relieving the distress by providing new cash and by streamlining operations. However, case studies reveal that personal financial benefits sometimes accrue to the nonprofit decision makers who agree to the conversion. In cases when the nonprofit management goals stay in place after conversion, the conversion may be motivated by the desire to gain greater access to capital; raising equity capital is possible for for-profits but violates the nondistribution constraint for nonprofits.[5]

The question of whether society gains a net benefit from these increasing conversions has come to interest health economists. There are at least two possible efficiency gains. Dynamic efficiency involves better access to capital enabling the rapid development of needed facilities, and efficiency of operation. Efficiency of operation may occur if new management and control eliminate entrenched practices and streamline operations. In addition, this type of efficiency difference should be observable in cost-efficiency studies that compare nonprofit hospitals with for-profits, a subject to which we return shortly.

The experience with conversion provides another opportunity to test the implications of ownership status. Picone, Chou, and Sloan (2002), Shen (2002), and Farsi (2004) all find that a short-term decline in quality measured by mortality rates usually follows conversion from nonprofit to for-profit status. Santerre and Vernon (2005) warn that conversion, by altering the nonprofit/for-profit mix in a market area, has implications for market-level economic efficiency. They suggest that the typical U.S. hospital market has an inefficiently large number of nonprofits.

The Relative Efficiency of Nonprofits versus For-Profits

Hospitals may be efficient in production but even so, they cannot control the other inputs into the community's health production, such as exercise and diets. Ellis (1993) contends that the combination of nonprofit hospital care and the community's other health production inputs will most probably produce community health inefficiently.

Are Nonprofit Health Care Firms Less Technically or Allocatively Efficient?—Hospital and Nursing Home Studies

Frontier studies examine hospital efficiency by attempting to identify the best possible practices. A firm is observed to be inefficient when it falls short of the best possible production practice frontier or above the cost frontier. Zuckerman et al. (1994), as well as Folland and Hofler (2001), found little or no difference between for-profit and nonprofit hospitals. A meta-analysis by Hollingsworth (2008), examining over 300 studies, led him to conclude cautiously that the public hospitals were somewhat more efficient than the other two organizational forms.

Nursing homes provide another test for nonprofit efficiency. Nonprofit homes sometimes appear to be less efficient merely because they are offering a higher quality of care (see Box 13.1 for a discussion). Nevertheless, Garavaglia et al. (2011), when adjusting for quality, found the for-profit homes to be more efficient. Santerre and Vernon (2007) contrast this by looking at industry-level efficiency. The two authors find that nursing home industry efficiency is higher when the mix of ownership types includes a larger share of nonprofit nursing homes.

BOX 13.1

Why Are Registered Nurses' Wages Higher in Nonprofit Nursing Homes?

Holtmann and Idson (1993) observed, as have others, the differentially higher wages that registered nurses receive in the nonprofit nursing home sector. They proposed that nonprofits pay nurses higher wages to get higher quality. Using econometric analyses, the authors discovered that in fact the differential wages reflect quality-enhancing characteristics of the nurses, such as years of experience and length of tenure on the current job, rather than differential rewards per level of experience or length of tenure.

Conclusions

This chapter has examined the nonprofit firm in the health care sector. We began with a description of the nonprofit firm, noting the importance of the nondistribution constraint. We then asked why nonprofit firms exist. We found two principal accounts. Weisbrod explained how nonprofit firms might arise to provide public goods that are neglected by the private markets and the government. Hansmann explained how nonprofit firms might reduce or eliminate a contract failure that arises because consumers may not trust the profit-motivated firm to perform faithfully certain functions, often charitable ones. Under these theories, the nonprofits can improve the well-being of the community, overcoming the for-profit firm's tendency to underproduce in the presence of beneficial externalities.

We then investigated three analytical models of nonprofit hospital behavior. The Newhouse hospital model examines the desire to provide service to the community, with the quality of care often being as important as the quantity. The Lakdawalla-Philipson model exploits a middle ground to explain the entry and exit behaviors of nonprofits. In contrast, the Pauly-Redisch hospital model involves physicians' control, used to maximize the average physician's income. We contrasted these two hospitals, which show the range of behaviors from purely altruistic concerns for the community to pure profit-maximizing interests. We showed the irony that, under pressure of competition, the differently motivated firms may behave nearly the same.

Finally, the nonprofit and for-profit health care firms were contrasted regarding efficiency. The data from efficiency studies show relatively little difference between the ownership types.

The first half of the twentieth century saw radical changes in the structure of the health care industry, particularly in the prevalence of nonprofit firms. Since then we have gained a better understanding of the economic factors that make nonprofits attractive, and we are beginning to understand what can make them increase or decrease in number.

Summary

1 The defining characteristic of a nonprofit firm is the nondistribution constraint. Furthermore, nonprofits are typically tax exempt, and donations to nonprofit firms receive favorable tax treatment.

2 The Weisbrod theory for the existence of nonprofits proposes that nonprofit firms arise to fulfill unmet demands for public goods.
3 The contract failure theory for the existence of nonprofits proposes that nonprofit firms are advantageous under circumstances where it is difficult or impossible for the purchaser of the good to verify the delivery and the quality of the good.
4 Altruistic conceptions of the nonprofit are exemplified by the Newhouse model, in which the hospital decision makers choose preferred combinations of quality and quantity of care subject to a break-even constraint.
5 A model where the nonprofit values both profit and altruistic service to the community illuminates entry and exit behavior.
6 The physicians' cooperative model depicts the hospital under de facto physician control exercised to maximize average physician income.
7 Evidence suggests that nonprofit hospitals in the United States are not very different in economic efficiency from for-profit hospitals.

Discussion Questions

1 What is the nondistribution constraint? In what way is the nondistribution constraint circumvented in the Pauly-Redisch model? What implication does this have for the efficiency of the Pauly-Redisch nonprofit hospital?
2 If an agency could cheaply, and accurately, monitor the delivery and quality of care by health care firms, would there be any contract failure in health care remaining? Would there be any need for nonprofits in health care? Would any arise?
3 In the Lakdawalla-Philipson model, why do the authors claim that the minimum average cost of the for-profit will determine the industry price?
4 Suppose that population growth expands the quality–quantity frontier of a Newhouse utility-maximizing nonprofit hospital. How would its choice of quantity and quality change? In your view, is a for-profit hospital likely to respond more quickly to population growth? Discuss.
5 Under which of the models of hospital behavior described in this chapter does the tax-exempt status of nonprofit hospitals make the most sense? Under which does it make the least sense?
6 Can we say which are the most efficient hospitals—nonprofits or for-profits? Which are the most efficient nursing homes? What qualifications apply to our present knowledge in each case? What is your view?
7 Why might information problems lead to consumer preferences for nonprofit provision of some goods and services? Reconcile your answer with the observation that most physician care, drug products, and many other services are provided by for-profit firms.
8 In what sense do nonprofits earn "profits" and need to earn "profits" to survive?
9 What are some cost advantages that nonprofits have over for-profits? Are there any disadvantages?
10 Explain the logic behind the argument that donations have characteristics of a public good.
11 Weisbrod and Hansmann present different theories on the existence of nonprofit organizations. Compare and contrast them in regard to the types of firms and the ways they are financed.

Exercises

1 In Figure 13.1, if two additional voters had demand curves equal to D_0, what amount of the public good would tend to be provided by the democratic government? Which voters would be unlikely to promote a nonprofit? Which would be the most likely?
2 In Figure 13.1, suppose that Voter 5 comes to value the public good even more than before. Will there be an increase in the amount provided through the median voter model of the voting process? Why or why not? Suppose that Voter 5 can bribe one of the other voters to change his or her preferences. Which one will Voter 5 approach?
3 Under the physicians' cooperative model, if the supply price of physicians were to rise, how would this affect the equilibrium staff size in the open-staff case? How would it affect the optimal staff size in the closed-staff case?
4 Of the Newhouse and the physicians' cooperative models, which nonprofit hospital is likely to produce more quantity and quality in equilibrium with barriers to entry? In long-run equilibrium, with free entry and exit?
5 Consider Figure 13.3. Are these long-run average cost curves (*LRAC*)? What accounts for the nonprofit advantage?

Notes

1 The history of hospitals and the relative importance of nonprofit versus for-profit status are further explored in Bays (1983), Frech (1990), and Temin (1988).
2 The essential idea was developed further in a model by Easley and O'Hara (1983).
3 For further study, see Cohen and Spector (1996) and Davis (1991).
4 Gulley and Santerre (1993) and Hansmann (1987) examine the effects of tax rates on nonprofit market shares.
5 For further investigation of conversions, see also Mark (1999), Cutler and Horwitz (1998), and Goddeeris and Weisbrod (1998).

Chapter 14

Hospitals and Long-Term Care

In this chapter

- Background and Overview of Hospitals
- Hospital Utilization and Costs
- Closures, Mergers, and Restructuring
- Quality of Care
- Nursing Homes
- Hospice, Home Health, and Informal Care
- Conclusions

In 2013, there were 5,686 hospitals with 915,000 beds in the United States (Table 14.1). Of the 18 million persons employed in health services industries, 4.8 million were employed at hospitals. Hospital care is the most visible component of total health care spending, and the hospital remains at the center of an evolving health economy. At the same time, the aging of the population has created a major challenge for the adequate provision of long-term care. This chapter provides an overview of the hospital and long-term care sectors. It also examines several controversies. For hospitals, these include the "medical arms race;" cost shifting; hospital quality; and the effects of managed care, the Affordable Care Act, and hospital consolidation. For nursing homes and long-term care, we examine various quality, demand, and cost issues, as well as the possible substitution of informal care for nursing home care.

Background and Overview of Hospitals

We distinguish among the many types of hospitals by using four criteria: length of stay, type, ownership, and size. Hospitals are categorized as short stay (usually less than 30 days) or long term (usually more than 30 days). The community hospital is the type with which the general public is most familiar. It consists of all nonfederal general hospitals that provide acute, short-term care.[1]

Many community hospitals are also teaching hospitals, with residency programs approved by the Accreditation Council for Graduate Medical Education. Other hospital types are

Table 14.1 Hospital Data

	1980	*1990*	*2000*	*2013*
All hospitals	6,965 (1,365)[a]	6,649 (1,213)	5,810 (984)	5,686 (915)
Federal hospitals	359 (117)	337 (98)	245 (53)	213 (39)
Nonfederal hospitals	6,606 (1,248)	6,312 (1,113)	5,565 (931)	5,473 (876)
Community	5,830 (988)	5,384 (927)	4,915 (824)	4,974 (796)
Nonprofit	3,322 (692)	3,191 (657)	3,003 (583)	2,904 (544)
For-profit	730 (87)	749 (102)	749 (110)	1,060 (135)
State–local government	1,778 (209)	1,578 (169)	1,163 (131)	1,010 (117)
Psychiatric and other long-term community hospitals	702 (256)	892 (183)	631 (105)	487 (79)
Occupancy rate[b]	75	67	64	63
Admissions (per 1,000 pop.)	159	125	117	106
Average length of stay (days)	7.6	7.2	5.8	5.4
Outpatient visits (per 1,000 pop.)	890	1,207	1,852	2,146

Notes: [a] Numbers in parentheses are beds in thousands.

[b] Percent of beds occupied. The latest value shown is for 2012.

Sources: U.S. Department of Commerce, *Statistical Abstract of the United States, 2016*: *Proquest* Online Edition, and earlier issues of the *Statistical Abstract*; and U.S. Department of Health and Human Services, *Health, United States* (2014 and earlier issues).

mental, including those treating alcoholism and other chemical dependencies; tuberculosis and other respiratory diseases; and other specialties (e.g., maternity, orthopedic, and rehabilitation).

Hospital ownership can be private or public (federal, state, county, or local). The former category consists of either nonprofit or proprietary (for-profit) hospitals. Table 14.1 indicates that there were 1,010 state and local short-stay hospitals in 2013. Nonprofits dominate the remainder.

Hospital size is generally measured by number of beds. This does not mean that one doubles hospital size by putting a second bed in each room, but rather that the support services, types of equipment, and to some extent administrative staff are related to the number of people that the hospital can house, and hence the number of beds.

Most short-stay hospitals are relatively small with fewer than 200 beds. However, hospitals that have more than 200 beds account for 68 percent of all beds. The largest hospitals usually are affiliated with university medical schools and provide tertiary care in addition to primary and secondary care. Tertiary care consists of the most complex procedures such as organ transplant surgery and open-heart procedures. The typical community hospital is limited to secondary care, consisting of the more common surgical and medical procedures. Primary care consists of the kinds of preventive and curative care received by patients who are not hospitalized.

History

Hospitals date back to ancient Egypt and Greece. Since then, places of healing in many countries were organized by religious establishments. Illness was closely associated with a lack of faith or superstition, and priests often administered care. Even today, it is not unusual for the afflicted to believe that they are being punished or cursed, and in some parts of the world, shamans and other "medicine men" are called upon to exorcise evil spirits.

Early hospitals in the United States were associated with the poor or with mental and infectious diseases, and medicine was practiced mainly at the home. This picture changed as more effective surgery became possible following scientific and technological advances in the last half of the nineteenth century. The modern U.S. hospital emerged at the turn of the twentieth century. In particular, important advances in antisepsis to help fight off infections greatly increased the probability of surgical success. Major advances in anesthesia, anatomy, and physiology, and the invention of the X-ray, also contributed.

Two nonscientific factors helped accelerate the process. One was the rapid pace of urbanization resulting from industrialization. Rural areas could not support sophisticated hospitals because of transportation problems and low population densities. Urbanization also created health problems, such as outbreaks of infectious disease that were much less common in rural areas and that required hospitalization.

The second factor was a financial one. Early hospitals relied on philanthropic contributions or state and local government funds. These alone would have been inadequate to support the growing numbers and costs associated with the modern hospital. Urbanization created wealth, and the rise of an urban middle class led to a greater ability to pay. Also supporting the hospitals was the innovation of third-party payment through private insurance and workers' compensation, originating in the early 1900s.

The opening of The Johns Hopkins Hospital in Baltimore in 1885 was a significant milestone. Though few other hospitals would ultimately be able to emulate or compete with

its international reputation, The Johns Hopkins Hospital introduced the latest advances in medical technology and hospital design. Through its affiliation with The Johns Hopkins University, it became a model of the teaching and research hospital.

Organization

The typical nonprofit community hospital is governed by a board of trustees that selects the president and approves most major decisions. Traditionally, board members were often leading members of the community known for their ability to contribute or raise funds. In recent years, the increasing financial complexity of decisions facing the modern hospital has made a strong business background an extremely important qualification.

The hospital's decision-making power rests with the medical staff rather than the administrators or the board. To appreciate the significance of this phenomenon, consider that the medical staff in most hospitals is composed of admitting physicians, who are not hospital employees. Physicians apply for staff privileges to admit patients and perform certain procedures. Because physicians often have admitting privileges at several hospitals and bring patients to the hospital, admitting physicians have considerable influence on hospital decision making by serving on many committees relating to hospital governance and patient care. As such, the hospital has been referred to as the physician's "rent-free workshop," where the physician can direct substantial resources for patient care but is not held directly accountable for those resources. Physicians also bill separately for their services.

To deal with the conflicts and cost pressures created by the traditional system, more hospitals now rely on permanent physician-employees who are paid salaries or combinations of salaries and bonuses, the latter driven by various incentives. As these staff physicians are no longer the driving force to admit patients, hospitals with permanent staff physicians must have other means to attract patients. The source of patients for such hospitals is often affiliation with or ownership of HMOs. Hospitals also advertise directly through the broadcast or print media or purchase physician practices to gain new patients.

The hospital industry has undergone major change. Due to declining inpatient utilization, many smaller hospitals have closed while others have merged or reorganized. Hospitals face considerable pressure to join networks of providers in order to participate in managed care plans and to become diversified health care centers with expanded primary care facilities. Many hospitals have concentrated resources on freestanding outpatient surgery units and other outpatient programs such as cardiac rehabilitation. Table 14.1 reveals the extent of these changes. Inpatient admissions per capita in community hospitals dropped 33 percent from 1980 to 2013, the average length of stay by 29 percent, and the occupancy rate dropped from 75 to 63 percent. Over the same period, the number of outpatient visits per capita more than doubled, increasing by 141 percent.

Despite this dramatic shift to outpatient care, hospitals will continue to face challenges, especially as Medicare and Medicaid budgets tighten. Governments account for about 38 percent of hospital revenues, although this figure can be far higher for many urban and small hospitals that rely heavily on poor and elderly patients. The ACA and other proposals to reform Medicare and Medicaid also tend to focus on reducing hospital spending and reimbursement rates. Thus, hospitals that disproportionately depend on Medicare and Medicaid are especially vulnerable. Nevertheless, in an increasingly competitive, cost-conscious environment, all hospitals are under considerable pressure to respond quickly to new incentives and opportunities.

Regulation and Accreditation

Hospitals are subject to a wide variety of state and federal regulations over quality, costs, and reimbursement. Hospitals are licensed at the state level, although licensure is often focused on the adequacy of the hospital's physical plant and other inputs. Hospitals have their own quality assurance programs but federal legislation established professional standards review organizations (PSROs) in 1971 to monitor quality while limiting utilization. After considerable controversy regarding their effectiveness, PSROs were replaced in 1984 by peer review organizations (PROs) that performed case-by-case peer review and monitored Medicare utilization in hospitals and other facilities. PROs were often dominated by physicians and hospitals and their impact was questioned. In the mid-1990s, the Centers for Medicare and Medicaid Services (CMS) adopted the position that case-by-case inspection brings only marginal improvements in quality. In 2002 PROs were replaced by quality improvement organizations (QIOs), intended to monitor and improve care.

Hospitals also are subject to numerous other regulations and requirements. Many of these relate to reimbursement, such as Medicare's prospective payments system (PPS) and various forms of state rate regulation. Certificate-of-Need (CON) laws limit capital spending, and hospitals are subject to antitrust laws intended to promote competition.

In addition to meeting licensure and regulatory requirements, most hospitals and many other health care facilities seek accreditation from the Joint Commission on the Accreditation of Healthcare Organizations (JCAHO). The JCAHO is a private, nonprofit organization that was founded in 1952 and has a board dominated by representatives from physician and hospital associations. It sets standards for patient safety and quality of care. Hospitals seeking accreditation are evaluated by a visitation team, which examines hospital compliance with JCAHO standards. To maintain accreditation, the hospital must undergo an on-site review every three years.

Many third-party payers reimburse only for care provided in accredited hospitals. Although hospitals can be evaluated by federal inspectors to qualify for Medicare and Medicaid reimbursement, JCAHO accreditation also satisfies the federal requirement. The JCAHO is clearly a powerful organization, and the potential for conflict between professional self-interests and public interests is evident. The influence of the JCAHO can be used to limit hospital competition and to protect physicians against other groups of providers, such as chiropractors and doctors of osteopathy, by denying them access to hospitals or influence within hospitals.

Hospital Utilization and Costs

The relentless growth of hospital costs has served as the impetus for many forms of regulation and other policy initiatives. Table 14.2 shows the increases in total hospital costs, as well as costs per day and costs per admission. Hospital costs account for 32 percent of national health expenditures and although they have decreased over the past two decades as a percentage of all health care costs, they have nonetheless increased at an annual rate of 9.2 percent since 1960. Table 14.2 also shows that a small and dwindling share has been paid out-of-pocket, especially after Medicare and Medicaid were introduced in 1965. These programs now represent 43 percent of all hospital costs.

Despite a growing and aging population, the hospital cost increases do not result from more inpatient admissions. The number of hospital beds has been declining for many years

(Table 14.1), and occupancy rates for community hospitals in 2013 are only slightly higher than the 62 percent bottom reached in 1997. Rather, the cost per day and cost per admission and the shift to outpatient services, leaving hospitals with high fixed costs, are the main driving forces. The influence of these determinants of hospital costs is intertwined closely with numerous features of health care markets including changes in technology and reimbursement methods. We develop these issues and many others relating to hospital costs and health care system reform in other chapters. In this section, we address two of the many other concerns relevant to the growth of hospital costs. The first concerns the effect of competition on costs, and the second deals with cost shifting.

Table 14.2 Hospital and Nursing Home Costs

	1960	*1970*	*1980*	*1990*	*2000*	*2013*
National health care spending ($ billions)[a]	27.4	74.9	256	724	1,377	2,919
Hospital care ($ billions)	9.0	27.2	101	250	416	937
Percent of total health spending	32.8	36.3	39.4	34.5	30.2	32.1
Percent of hospital care paid by						
Out-of-pocket	20.6	9.0	5.4	4.5	3.2	3.5
Private health insurance	35.6	32.5	36.7	38.7	34.1	37.1
Medicare and Medicaid	0	29.4	35.3	37.5	46.8	43.4
Medicare	—	19.7	26.1	26.9	29.7	25.9
Medicaid	—	9.7	9.2	10.6	17.1	17.5
Other government and private[b]	43.8	29.1	22.6	19.3	15.9	15.9
Average hospital cost ($)						
Per day	—	—	245	687	1,149	2,157
Per stay	—	—	1,851	4,947	6,649	11,651
Nursing home care ($ billions)	0.8	4.0	15.3	44.9	85.1	155.8
Percent of total health spending	2.9	5.3	6.0	6.2	6.2	5.3
Percent of nursing home paid by						
Out-of-pocket	74.7	49.5	40.7	40.3	31.9	29.4
Private health insurance	0	0.2	1.3	6.2	8.8	8.1
Medicare and Medicaid	__	26.8	48.2	40.4	50.1	52.3
Medicare	—	3.5	2.0	3.8	12.7	22.2
Medicaid	—	23.3	46.2	36.6	37.4	30.1
Other government and private	25.3	23.5	9.8	13.1	9.8	10.3

Notes: [a] All monetary values are in nominal dollars.

[b] Includes the Department of Veterans Affairs, CHIP, other state and local spending, workers' compensation, and other public and private sources.

Sources: U.S. Department of Health and Human Services, *Health, United States*, *2014*; and U.S. Department of Commerce, *Statistical Abstract of the United States* (2016 and earlier issues).

Competition and Costs

Consumers generally welcome increased competition as a driver of lower prices, greater availability, and improved innovation and quality. Whether increased competition among hospitals provides similar benefits is of considerable policy and academic interest. Such concerns are interesting and unique. With the exception of natural monopolies resulting from economies of scale, economists usually endorse competition as being in the best interest of consumers. Evidence is substantial that higher levels of seller concentration in most markets lead to higher prices and reduced choices. Indeed, this is the premise behind federal and state antitrust laws.

Some analysts suggest that the hospital market is an exception to the standard paradigm. They argue that hospital competition has encouraged an unproductive and costly medical arms race (MAR), as described in Box 14.1, with unnecessary duplication of expensive capital equipment as well as unnecessary expenditures on advertising in order to attract patients. Competition also may create pressure to fill beds through questionable admissions.

Why should hospital care be different? The answer lies with the reimbursement mechanisms traditionally used by insurers, which paid hospitals on a retrospective cost basis. Higher costs generally meant higher payments to hospitals. Unlike other industries, where sellers must compete on the basis of price for customers, retrospective reimbursement meant that hospitals were largely immune from the discipline exerted by the competitive process.

BOX 14.1

Game Theory and the Medical Arms Race (MAR)

The 2002 Academy Award-winning movie *A Beautiful Mind* brought considerable public attention to John Nash and his contributions to game theory. Game theory is a powerful analytical tool used increasingly in economics and many other disciplines. It can be used, for example, to show why it may be in the best interests for each hospital to engage in a MAR even when hospitals as a whole are negatively affected. Game theory begins with a payoff matrix of the type shown below. Suppose there are two large hospitals, A and B, in a market, each facing the decision of whether to add an expensive heart transplant unit without knowing what its rival will do. The payoff matrix shows the total profit for each hospital (with values for A's profit shown first) resulting from the four combinations of strategies. For example, if both adopt (the "northwest" cell), each hospital will have a total annual profit of $100 (million). If A alone adopts (the "northeast" cell), assume that it will have a significant advantage resulting in a profit of $200 (million), while B loses $50 (million).

Game theory tries to predict a solution, that is, the strategy chosen by each participant. It is clear that both hospitals with a combined profit of $300 (million) will be better off if neither introduces the unit. However, if the hospitals cannot agree (e.g., they may not trust each other or they may believe that antitrust laws preclude cooperation), game theory predicts a solution in which each hospital will adopt the unit and combined profits will be $200 (million). Why? Given the payoff matrix, each hospital has a dominant strategy. That is, regardless of what Hospital B does, A will always have a higher profit by adopting rather than not adopting, that is, $100 (million) versus –$50 (million) if B adopts and $200 (million) versus $150 (million) if B does

not adopt. Similarly B's dominant strategy is to adopt and, hence, a scenario results consistent with the MAR hypothesis.

Students of game theory will recognize this as an example of the prisoner's dilemma and the solution as a Nash equilibrium. McKay (1994) and Calem and Rizzo (1995) provide other applications of game theory to other decisions including hospital quality and specialty mix. In addition to decisions involving the acquisition of technology and introduction of new services, game theory can provide insight into hospital advertising and other forms of nonprice competition.

		Hospital B	
		Adopt	*Do Not Adopt*
Hospital A	Adopt	100, 100	200, –50
	Do Not Adopt	–50, 200	150, 150

This situation has greatly changed over the past two decades. Hospitals as well as insurance companies must compete for their managed care business through price and quality. Hospitals also are now reimbursed by many major third-party payers on a prospective basis at rates that are independent of their actual costs. It would thus appear that hospitals have a strong financial stake in being efficient and in avoiding capital investments that are not profitable.

Kessler and McClellan (2000) examined the effects of hospital competition on the costs and outcomes for Medicare beneficiaries who incurred heart attacks. They found that, prior to 1991, competition improved outcomes in some cases, but also raised costs. After 1990, there were substantial decreases in costs and substantial improvements in outcomes. A part of the welfare improvement resulted from the higher HMO enrollments over this period.

Elsewhere, Zwanziger and Mooney (2005) studied HMOs in New York State which until 1996 regulated the rates (determined largely by historical costs) private insurers were required to pay for inpatient care. After the 1996 reforms, HMOs were able to negotiate lower prices with hospitals that were located in more competitive markets.

Finally, we observe other effects of managed care, including its role in the large decline in the number of hospitals and beds since 1980. Dranove and colleagues (2002) describe the financial pressures created by managed care for hospital consolidation and improved efficiency. Their work demonstrates the substantial impact associated with the growth of managed care. For the average market, the consolidation between 1981 and 1994 attributable to managed care represented the equivalent of a decrease to 6.5 equal-sized hospitals from 10.4 such hospitals.

A second study by Dranove and colleagues (2008) examined whether the "managed care backlash" and consequent easing of restrictions on patient choices beginning in the mid-1990s affected the relationship between hospital prices and hospital concentration. MCOs should be able to extract greater price concessions in more competitive hospital markets especially when their members are more willing to be directed to selective hospitals. In the limiting case of a single hospital in a market, the MCO will have no bargaining ability. The research results show that the price–concentration relationship grew stronger between 1995 and 2001, but that it peaked in 2001, and possibly even reversed after 2001. The authors conclude (p. 374) that despite growing concentration, there has not been a "collapse in the

price–concentration relationship. MCOs continue to appear to be playing competitive hospitals off against each other to secure discounts, though with possibly less effectiveness than in the peak year of 2001."

HOSPITAL COST SHIFTING For various legal and ethical reasons, hospitals provide substantial amounts of uncompensated care. Most of this care is provided to uninsured indigents, but uncollectibles from incompletely insured patients are also considerable. In addition, many third-party payers place stringent limits on reimbursement rates, and proposals to reduce Medicare and Medicaid expenditures typically call for further reductions. After an initial period of generous payments under PPS, by 1993 the payment-to-cost ratios for both Medicare and Medicaid were only 90 percent, compared to 130 percent for private patients. The ratios for Medicare and Medicaid subsequently increased but dropped back to the 1993 levels by 2004.[2] They have remained at about 90 percent compared to a private payer rate of 144 percent in 2013.

Are the costs of uncompensated care and "discounts" to some third-party payers passed on by hospitals to other patients as is often claimed? If Medicare and Medicaid cuts are passed on to others, there would be no savings to society but merely a shifting of the hospital cost burden. Similar shifting would occur if the number of uninsured or poorly insured increases because of an increase in part-time employment in the services sector and cutbacks in fringe benefits by some employers.

Intuition suggests that these costs are shifted. After all, services must be paid for and it stands to reason that the burden for nonpayers must be picked up by others. However, the issue may not be as simple as it first appears. To see why, we develop a model of hospital fee determination.

We examined a variety of hospital behavior models in Chapter 13, but for simplicity assume that hospitals maximize profits (or the undistributed residual in the case of non-profits). Suppose also for simplicity that there are just two groups of patients: private (insured or self-pay) and Medicare. The downward-sloping demand curve for the private sector and the constant Medicare hospital reimbursement rate (R_1) per patient are shown in Figure 14.1. The private demand curve (panel A) is negatively sloped because at least some patients economize or substitute other services as their out-of-pocket obligations increase, and hospitals that raise fees lose patients to other hospitals. Assume further that R_1 covers the average variable cost (C_1) for a fixed number of Medicare patients seeking admission (i.e., Q_2 Medicare patients in panel B) but that the rate does not necessarily cover all costs. Finally, assume that the hospital is operating below capacity, as is the case for many hospitals, and that C_1 is constant over the relevant range and equal to marginal cost.

If the hospital is unable to price discriminate in the private sector, it will accept Q_1 private patients (the quantity where marginal revenue equals marginal cost) and charge the price, P_1. It also will accept all Medicare patients (Q_2) at the Medicare rate[3] so that the hospital treats a total of ($Q_1 + Q_2$) patients. Total revenues of ($P_1Q_1 + R_1Q_2$) produce a surplus over variable costs equal to $(P_1 - C_1)\ Q_1 + (R_1 - C_1)\ Q_2$.

Suppose that the Medicare reimbursement rate is lowered to R_2, but that the rate still covers the average variable and marginal costs so that the hospital continues to accept Q_2 Medicare patients. Contrary to intuition, it makes no sense to increase prices in the private sector. Hospital surpluses diminish at prices above P_1 because the hospital will lose private sector patients whose marginal revenue exceeds marginal cost. The optimal private rate remains at P_1 for the Q_1 private patients, and no cost shifting occurs.[4]

This is not the end of the story because the hospital's revenues are reduced. In the long run, revenues must cover all costs, or else the hospital cannot survive. If revenues exceed costs for the hospital in Figure 14.1 after the lower Medicare rate, then the private rate may

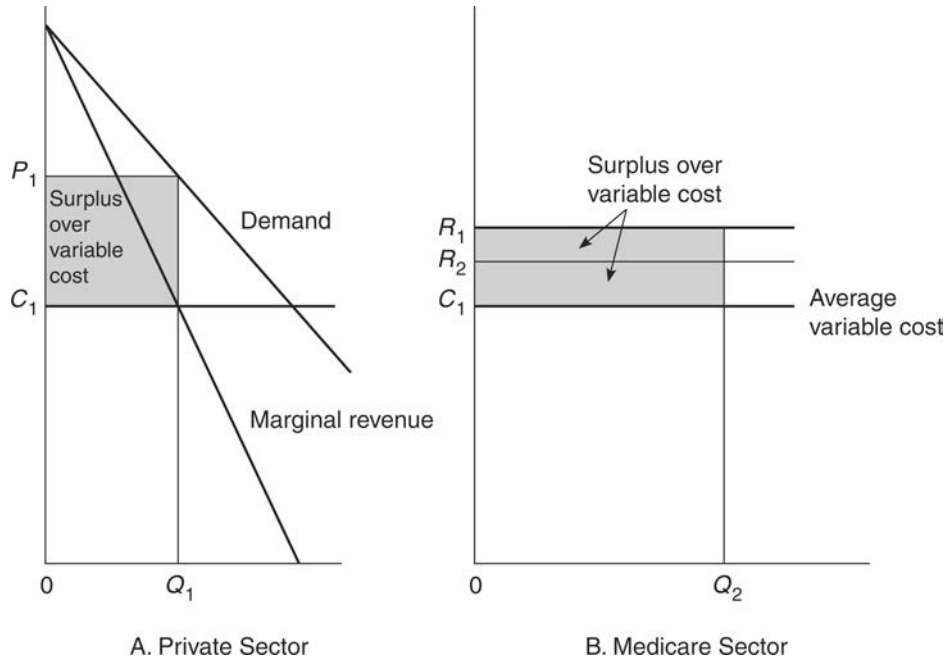


Figure 14.1 Analysis of Changes in the Medicare Hospital Reimbursement Rate

be unaffected in the long run as well. Suppose, however, that the hospital's revenues now fall short of total costs. If it cannot reduce costs, the hospital and others in the same situation may be forced out of business or forced to merge. As this happens, demand will increase for the remaining hospitals and, as a result, the private rate also could rise.

Determining the burden of lower public reimbursement is complex. The burden ultimately can be borne by many groups through reductions in the number of hospitals, lower compensation to hospital employees as the demand for their services diminishes, reduced access to care for those with public insurance or those receiving uncompensated care, and higher fees to the private paying groups.

A review of the evidence on cost shifting by Morrisey (1995) indicates that cost shifting through higher prices has taken place but that it is far from complete. One study included in his review shows that California hospitals reduced the amount of uncompensated care by 53 cents for every $1 decrease in their discounts to third parties. This would have been unnecessary if the hospitals could have shifted the costs to others.

More recently, following reductions in Medicare payments to hospitals, Wu (2010) found relatively little cost shifting overall but large variations across hospitals. Those where Medicare reimbursement was small relative to private insurance were able to shift nearly 40 percent of the Medicare cuts. Hospitals that relied more heavily on Medicare patients were much more limited in shifting costs.

Finally, Robinson (2011) also examined the effects of Medicare payment shortfalls by studying revenue–cost margins for seven specific conditions such as knee and hip replacement. Hospitals in concentrated markets are more likely to raise prices for privately insured

patients, i.e., shift costs, while those in more competitive markets tend to increase efficiency by reducing costs. This distinction is of increasing importance because, as described below, the rate of growth of payments to hospitals under the ACA is being reduced. Policymakers need to have greater certainty in the consequences of these payment reductions.

Closures, Mergers, and Restructuring

We have already referred to some of the dramatic changes affecting hospitals. The growth of managed care and the introduction of reimbursement methods that discourage inpatient care and long lengths of stay have contributed to declining inpatient utilization. In response, capacity has been reduced through the sharp drop since 1980 in the number of hospitals and beds (Table 14.1). Even so, with occupancy running at just 64 percent in 2013, excess capacity remains one of the most visible and significant characteristics of the hospital industry.

Although hospital closures can be painful to a community, the restructuring of the hospital industry should be viewed as a market response to cost-containment efforts. Nevertheless, it remains important to determine just how well the market works for this sector, and, in particular, whether inefficient hospitals are more likely to close.

Cleverly (1993) examined 160 community hospitals that closed between 1989 and 1991. Most were small, located in rural areas, and had sustained progressively larger losses for several years before closing. High costs and high prices, low utilization, and little investment in new technology were common features. From the characteristics of failed hospitals, Cleverly describes the road to failure. High prices and lack of investment in technology drive patients away. With lower utilization, costs per patient increase and cash flows become negative. The deteriorating liquidity ultimately leads to closure.

The relatively large number of small, rural hospital closings has challenged policymakers to maintain access for rural populations. To prepare for unexpected influxes of patients, small hospitals have higher rates of excess capacity and, hence, lower occupancy rates than larger hospitals. Various federal programs provide subsidies to these hospitals. Nevertheless, rural hospitals can increase their chances of survival by practicing good management and responding to competitive pressures. Succi and colleagues (1997) found that rural hospitals gain an advantage and reduce the threat of competition by differentiating their services. Those that offer more basic services and high-tech services compared to the market average are less likely to close.

With the continued growth of managed care in the 1990s and hospitals' increased reliance on managed care patients, the pace of hospital restructuring accelerated. Nearly every hospital was facing serious financial and competitive challenges. Hospitals responded by merging, participating in multihospital systems, and by forming various strategic alliances with other hospitals and physician practices. There are two intended effects of these activities. First, by downsizing administrative units, eliminating duplication, and taking advantage of economies of scale through integration, hospitals seek to become more efficient and, therefore, more successful in competing for managed care contracts. Jantzen and Loubeau (2000) found that price and hospital participation in networks are very important to managed care organizations in awarding contracts. Second, hospitals and hospital systems, through their size and partnerships, seek to counter the pricing pressure and other demands that have been placed on them by managed care organizations.

The passage of the ACA in 2010 has only accelerated the merger frenzy. To help cover the costs of coverage expansions under the ACA, Medicare is reducing the payment updates to

hospitals for both inpatient and outpatient care. The amounts are substantial, on the order of $100 billion over a 10-year period. Other ACA changes, directed more at improving quality, can have significant financial consequences. For example, Medicare payments were cut by 1 percent in 2015 (about $370 million in total) to hospitals that had excessive "hospital acquired conditions" such as patient falls or leaving a foreign object inside a patient after surgery. Payments are also reduced to acute care hospitals that have excessive readmissions rates.

Even prior to the ACA, health economists sought a better understanding of the two principal effects of the restructuring: whether hospital care was produced at lower costs than would otherwise have been observed, and whether prices rose as a result of less competition among hospitals. For example, Harrison's (2007) examination of closures and mergers using comprehensive national data covering the period 1981–1998 found that increased market power, rather than improved efficiency, is the principal driving force for consolidations. Consistent with these findings, Melnick and Keeler (2007) showed that hospitals that were members of multihospital systems increased their prices between 1999 and 2003 at much higher rates than nonmembers.

The ACA has intensified research interest in the effects of hospital consolidation. A literature review by Gaynor and Town (2012) concludes that "increases in hospital market concentration lead to increases in the price of hospital care" and that "mergers in concentrated markets generally lead to significant price increases" (pp. 1–2). Surprisingly, the authors also conclude that "at least for some procedures, hospital concentration reduces quality" (p. 3). More recently, Dafney and colleagues (2016) found that even mergers among hospitals serving different markets can lead to higher prices.

Because these effects are undesirable from society's perspective, regulators and policymakers must be up to the challenge of assessing the restructuring efforts especially in light of evidence showing that there are cases where hospital closures on balance increase economic welfare (Capps et al., 2010) and that hospital closures do not adversely impact mortality rates or rates of hospitalization for populations in the affected hospital markets (Joynt et al., 2015).

BOX 14.2

Hospitals and Airlines: What Are the Lessons?

A provocative article, "Could U.S. Hospitals Go the Way of U.S. Airlines?," raises important and troubling questions about the potential adverse consequences of downsizing in the hospital sector. Altman and colleagues (2006) draw interesting parallels between the hospital and airlines industries, most notably through their historical lack of price transparency, limited competition, and cross-subsidies. With increased price transparency and competition from specialized low-cost airlines, the legacy airlines downsized, merged, cut unprofitable routes and capacity, and reined in wages and other costs. Despite these efforts, their financial state remains precarious.

What might happen if the hospital industry faces similar pressures? There are already strong efforts to increase price transparency through posting of prices on the Internet and other mechanisms. (See Reinhardt (2006) for his description of hospital pricing as "chaos behind a veil of secrecy.") Specialized clinics, ambulatory surgery

centers, and other freestanding outpatient facilities are increasingly competing with the general hospital. The Altman article suggests that hospitals might be forced to follow the airlines' example by paring back capacity, services, staff, and the quality of care. Unprofitable patients and hospital units could be the first to go. High-cost communities that depend heavily on Medicaid and Medicare patients would be especially vulnerable if cross-subsidies were to be reduced.

The tight government budgets and cutbacks in employer-provided insurance that are likely to continue in coming years would appear to make prospects for the hospital sector even more dire. However, actual developments in the health care industries often surprise analysts. For example, Courtemanche and Plotzke (2010) show that the growth of ambulatory surgical centers, a seeming threat to hospitals, has had little effect on their outpatient surgical volume and no effect on their inpatient volume. Similarly, Bates and Santerre (2008) found that managed care organizations, another potential threat to hospitals, have not usually taken advantage of their monopsony power.

Quality of Care

In discussing asymmetric information in Chapter 10, we have already described how quality of care has emerged as a national priority. Hospital quality is often understood in two ways. Chapter 13 introduced theoretical objective functions for hospitals that include quality, where quality is represented in a broad sense through characteristics such as attractiveness of facilities, and the expertise and prestige of staff. The availability of high-tech units and services also falls into this category. However, quality can also be understood in terms of hospital mortality and error rates, readmission rates, and the rates at which a hospital meets established treatment processes and protocols. Much of the current concern with quality, and efforts to improve it, revolves around the latter sorts of performance indicators. Through a variety of voluntary and mandated mechanisms, hospitals are constantly being evaluated on the premise that closer scrutiny and publicly available report cards will encourage quality improvements. Patients can now compare hospital performance measures for specific conditions and procedures under the Centers for Medicare and Medicaid Services "Hospital Compare" initiative (www.medicare.gov/hospitalcompare/search.html?). If embarrassment about poor performance is not a sufficient motivator, hospitals have a major financial stake in raising quality to secure managed care contracts and to attract the growing numbers of patients that are choosing consumer-directed health plans.

Two reports published in the July 21, 2005, *New England Journal of Medicine* provide considerable insight into the quality of hospital care and the impact of public reporting. One (Williams et al., 2005) evaluated an initiative implemented by JCAHO in 2002 that required most accredited hospitals to report 18 standardized performance indicators for several common conditions. Seventeen indicators assessed processes of care; one was mortality. Analysis of quarterly data for heart attacks, congestive heart failure, and pneumonia indicated substantial gains in 15 of the 18 measures over the two-year study period. The gains were greatest for hospitals that had been the worst performers at the start of the evaluation period.

A second report (Jha et al., 2005) examined ten quality indicators for patients discharged at 3,558 hospitals in 2004, also for heart attacks, congestive heart failure, and pneumonia. Mean performance scores (representing proportions of patients who satisfied the criteria) were 89 percent for heart attacks, 81 percent for congestive heart failure, and 71 percent for

pneumonia. Performance varied substantially among hospitals, and many failed to provide patients with broadly accepted treatments at surprisingly high rates. It is possible that the scores reflect, in part, record-keeping omissions rather than failure to provide treatment, but such omissions themselves would indicate a quality problem.

Regional variations are also startling. Among the 40 largest hospital regions, Boston led the way in treating heart attack, with a performance score of 95 percent. San Bernadino was at the bottom with 83 percent. Overall, hospitals in the Midwest and Northeast outperformed those in the South and West.

Hospital quality has not been ignored in economic analyses and Chapter 10 describes several quality-related issues. Here we close our discussion with a revealing study by Tay (2003) that examined the impact of quality and distance on (non-HMO) Medicare patients admitted to a hospital for heart attacks. Medicare patients who are not enrolled in managed care plans have a wide range of hospital choices, and price is not an issue to them. Consistent with economic theory, Tay found that distance has a significant negative impact on demand, while quality, measured by various input and health outcomes indicators, has a significant positive impact. Remarkably, some patients are willing to travel much further for higher quality, even for conditions where a small delay to treatment can have a big impact on outcomes.

The bottom line is that quality matters a great deal to patients. It is also of increasing importance to third-party payers. We have previously described some ACA features that link hospital payment to quality. Financial incentives to improve quality may in part be driving hospital consolidation. Larger hospital systems may be able to better coordinate care and invest in technology and information systems that produce better patient outcomes. The challenge for the hospital is to restructure in ways that actually deliver quality improvements.

Nursing Homes

The rapid growth of the elderly population in many countries has led to considerable interest in the problems associated with long-term care. Long-term care encompasses a wide variety of services and arrangements used to care for the elderly and others with serious functional impairments.

In this section, we concentrate on the nursing home. Numerous economic issues have been investigated in the nursing home literature (Norton, 2000). Following some background information, we will focus on those involving quality, cost shifting, and financing the care.

Background and Costs

Traditionally, the elderly were cared for until death by their families so the nursing home is a relatively recent phenomenon. Raffel and colleagues (2002) trace its origins. The first "nursing homes" in the United States were the county poorhouses of the eighteenth and nineteenth centuries, established for invalids and those without families. Most of the patients were elderly, and the conditions were dreadful. Other state and local facilities evolved and some, usually with higher standards, were sponsored by religious and fraternal groups. The Social Security Act of 1935 provided funds for patients in private nursing homes, but the major increase in funding and nursing industry growth came after the 1965 passage of Medicare and Medicaid.

In 1960, nursing home care amounted to less than $1 billion, with 78 percent coming out-of-pocket (Table 14.2). Twenty years later, in 1980, spending reached $18 billion, with

Medicaid's share at 50 percent and only 40 percent coming out-of-pocket. By 2013, spending had grown to $156 billion, with Medicaid contributing 30 percent (another 22 percent came from Medicare) and patients or their families paying just 29 percent. Between 1963 and 2000, the nursing home population over age 65 grew from 446,000 to nearly 1.5 million. Since 2000, it has leveled off. In 2013, about 1.4 million residents occupied 15,700 nursing homes. About two-thirds are living in for-profit facilities.

The burgeoning nursing home population and the growth of costs are connected closely to Medicare and Medicaid. Medicare typically covers beneficiaries who are discharged from hospitals and require skilled nursing care to help recover from an acute illness. A skilled nursing facility provides round-the-clock nursing care and other medical supervision. Historically, Medicare reimbursed on a cost basis, but the Balanced Budget Act of 1997 mandated a shift to prospective reimbursement with fixed per diem payments determined by the category in which a patient is placed. The goal of this change was to shift the financial risk to nursing homes.

Medicaid, in contrast, pays for the long-term care of the poor including the nonelderly as well as elderly. It covers both skilled nursing care and intermediate (custodial) care. Because Medicaid is administered by the states subject to federal requirements, eligibility requirements and payment methods can vary widely. In the 1980s, many states phased in prospective payment systems that distinguished only between patients in skilled and intermediate care facilities. The rates were often set too low for heavy-care patients, discouraging nursing homes from admitting them. To provide better incentives, states increasingly are adopting case-mix reimbursement systems in which payment depends on a more extensive classification of patient types.

A 1987 legislative change created new standards that drove costs higher. All nursing homes participating in federal programs, including Medicaid-only facilities, must now meet the same standards as Medicare's skilled nursing facilities. Also, nursing homes participating in federal programs must evaluate each resident's needs and "provide services and activities to attain or maintain the 'highest practicable level' of function (physical, mental, and psychosocial well-being)." These requirements led to large increases in staffing and the use of rehabilitation services.

Quality of Care

Despite increasingly stringent state and federal requirements, the public cares most about quality. To many, nursing homes evoke powerful images of neglected and poorly treated patients. It is thus natural that health economists and other scholars have concentrated their efforts on nursing home quality.

As we have seen from those chapters that deal with quality issues, researchers must find ways to define and measure quality. Regulatory standards for nursing homes tend to focus on structure measured by the quantity and categories of inputs used to provide care, rather than evaluating how the care is delivered (process) or monitoring outcome indicators such as patients' satisfaction or their quality of life.

It follows that the most comprehensive and readily available information on nursing homes and other health care institutions comes from input data—their types and costs. Scholars have used these data to examine the relationship among quality and nursing home size, ownership, expenditures, and source of payment.

One would expect a positive association between size and quality as a result of economies of scale and scope. If such economies occur over some range of output, larger institutions would be able to provide the same quality of care at a lower cost, or a higher quality of care

for a given level of spending per patient. Davis's (1991) review of a large number of studies, including those that used process and structure measures of quality, suggests that no clear relationship exists.

A more important issue involves quality and cost or expenditure per patient. Do "you get what you pay for" in nursing care? Using structure measures of quality, not surprisingly, most analysts find a positive relationship, but these results are not meaningful. If inputs represent quality, one expects a close relationship between various measures of input and quality. Of the 18 process and outcome studies included in Davis's literature review, only 6 indicate a positive relationship between quality and cost or inputs, while the results in 11 are insignificant (one is negative).

Subsequent research reached similar conclusions. In one of the largest studies of long-term outcomes, Porell and colleagues (1998) found little relationship between quality and facility characteristics, such as size and ownership type. They even determined that higher RN staffing does not improve outcomes. Except for nursing homes at the highest staffing levels, where the top 10 percent were significantly better, Schnelle et al. (2004) also found few quality differences.

The failure to find positive relationships consistently is troubling. It indicates that improvements are needed in measuring quality as well as in formulating the statistical models used to estimate the relationships.

A third area of interest is in the relationship between type of ownership and quality. The previous chapter discussed the concept of contract failure within the context of nursing homes. Put simply, contract failure arises when quality is not easily observable. In the case of nursing homes, patients may believe that nonprofit organizations are more likely to serve their interests than ones motivated by profits. Is this view justified by the evidence? Nursing home costs per patient are higher for nonprofits (which have higher proportions of private-pay patients) so that structure measures are clear on this point. However, because analysts have not been able to detect an unambiguous positive relationship between quality and costs, it follows that they would have great difficulty in detecting any relationship between type of ownership and process or outcome measures of quality. Clearly, the relationship between quality and ownership is a complicated one, where profit status possibly plays a secondary role relative to other factors (Decker 2008).

Finally, Davis also reviews the literature on quality and the proportion of public-pay (Medicaid) patients. Many believe that nursing homes dominated by Medicaid patients are inferior. Expenditures per resident are lower in homes with higher proportions of Medicaid patients so that structure measures unequivocally support a negative relationship between quality and the proportion of Medicaid residents in a nursing home. Troyer (2004) found that Medicaid resident mortality rates were 4.2 percent and 7.8 percent higher than those for private-pay residents after one and two years, respectively. However, these differences declined when the analysis included controls for resident, market, and facility characteristics. It appears that, while Medicaid patients may be concentrated in lower quality nursing homes, the care given to patients does not vary by payment source.

Excess Demand

For many years nursing home observers have characterized the nursing home industry as having excess demand, and have argued that excess demand is one of the reasons for the allegedly inferior quality of care provided to public-pay patients. Economists are naturally intrigued by, and at the same time skeptical of, claims of persistent shortages of any commodity. Put simply, they believe that prices, and ultimately supply, will increase to eliminate the excess demand.

To examine the possibility of excess demand for nursing home care, Figure 14.2 introduces a variant of an approach originally developed by Scanlon (1980). It shows the demand and cost conditions for a representative nursing home. The demand curve reflects only the private demand (self-pay or insured), while R_1 represents the Medicaid reimbursement rate. The segment AC along R_1 shows the number of Medicaid patients seeking admission. For simplicity, assume a constant (horizontal) marginal and average variable cost (C_1) up to the capacity level (Q_c patients) where it becomes vertical (no more patients could be served at any cost).

Under the conditions represented here, the profit-maximizing nursing home will first select all private patients whose marginal revenue exceeds R_1 and then fill the remainder of beds with Medicaid patients. The nursing home admits Q_1 private patients paying a price, P_1, and ($Q_c - Q_1$) Medicaid patients, leaving an excess demand of BC Medicaid patients.

The shortage can be reduced or even eliminated by raising the Medicaid rate. At R_2, nursing homes will raise the private fee to P_2 and substitute $A'D'$ Medicaid for AD private patients. The excess demand is reduced to $B'C'$ from BC.[5] Conversely, reductions in the Medicaid rate will lower Medicaid admissions and the private fee but increase the excess demand. Similarly, an increase in private demand resulting from higher incomes or more prevalent private long-term insurance will reduce Medicaid admissions and increase the excess demand.

One additional aspect should be considered. If the nursing home industry is profitable, and evidence suggests that it is, one would expect entry of additional nursing homes to reduce the excess demand. However, nursing homes also are subject to Certificate-of-Need (CON) regulations, and it has been argued that legislators have intentionally used CON and other restrictions to limit nursing home entry as a way of limiting Medicaid spending.[6] There is a similar theme of legislative aversion to policies that would raise the Medicaid rate, which, as shown previously, increases public spending while squeezing out private patients.

Various tests have been proposed for the shortage hypothesis. For example, in our model, changes in reimbursement rates do not affect total utilization, just the composition

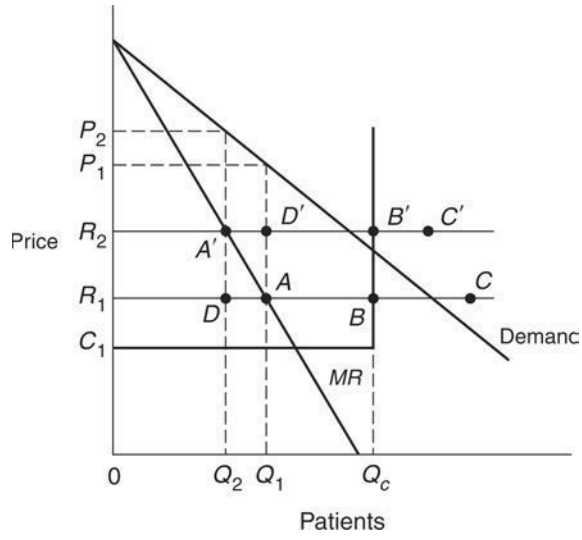


Figure 14.2 Analysis of Changes in the Medicaid Nursing Home Reimbursement Rate

between private and Medicaid patients. Any new beds will more likely be filled by Medicaid than by private patients because that is the population for which there is excess demand. Using national data for 1969 and 1973, Scanlon's empirical tests indicate considerable excess demand for Medicaid patients. However, states may vary widely in their willingness to fund public patients, and changes may have taken place in more recent years to reduce the shortages. For example, Nyman's (1993) estimates for 1988 for Minnesota, Oregon, and Wisconsin generally do not support the shortage hypothesis. The 2013 national nursing home occupancy rate of 81 percent also does not support suggestions of current shortages.

Financing Long-Term Care

The continuous growth of the population that will need long-term care (LTC); the requirement that patients must meet income, asset, and home equity tests to qualify for nursing home benefits under Medicaid; and the budgetary problems created by the growth of Medicaid spending have led to many proposals to reform Medicaid. The need to deplete one's assets is especially burdensome to the middle class. The Deficit Reduction Act of 2005 even extended the "look-back" period, where Medicaid examines the recipient's financial records, from three to five years. As a result, a variety of proposals have been introduced to help resolve this contentious issue. These range from proposals allowing individuals to have higher incomes and retain a higher proportion of their wealth to qualify for public assistance, to those that would cover everyone who meets certain medical requirements.

Federal policy to contain public spending centers on two strategies: (1) encourage home care and other potentially less costly substitutes for nursing home care, and (2) encourage more private coverage for LTC. Legislation passed in 1996 allows employers 100 percent deductions on their contributions to group plans. The legislation also provides some tax relief to individuals who itemize their returns and purchase tax-qualified policies. Nevertheless, according to the Kaiser Family Foundation, only 4.8 million had LTC coverage in 2014. Lapse rates are substantial (Konetzca and Luo (2010) provide a detailed analysis), and 2009 marked the first year in which the total number of existing LTC policies did not increase. Since then, many insurers dropped out of the LTC market and those that remain have increased premiums substantially.

What are the barriers to growth of private LTC insurance? They include confusion over the benefits provided by Medicare and private health plans, perceptions that such insurance is unaffordable or a "bad investment" if they are not going to use it (Gottlieb and Mitchell, 2015), and perceptions that governments will somehow meet LTC needs. The low probability (about 1 percent) that employed workers will need long-term care before they retire, coupled with uncertainty over future premiums and medical technology, increases the difficulty of selling LTC insurance to the working population. To overcome these barriers, some analysts propose additional tax incentives by allowing all taxpayers a 100 percent tax deduction for premiums or allowing tax-deductible LTC accounts, similar to retirement accounts. The Medicare drug legislation, effective in 2004, allows individuals to create health savings accounts that could be used to pay for qualified LTC premiums.

The Partnership Long-Term Care Insurance Program (PLTC) is another recent innovation under which states partner with private insurers to allow those with this insurance to keep additional assets as they spend down to qualify for Medicaid coverage. The goal of the program is to encourage more purchases of LTC insurance by those in the middle class. However, Lin and Prince (2013) found relatively few takers under this program and those who did came largely from those with greater wealth.

The more visible strategy to contain public spending is the federal government's emphasis on alternatives to nursing home care. The proportion of the elderly found in nursing homes has been decreasing. The elderly population grew nearly 18 percent between 1985 and 1995 and the number of nursing home residents also increased, but there was a striking 8 percent drop in their use rate. Bishop (1999) estimated that had the 1995 nursing home share remained at the 1985 level, a quarter of a million additional elders would have been placed in nursing homes. Where have all these patients gone? The high rates of growth of home health care and other LTC arrangements may provide an answer.

Hospice, Home Health, and Informal Care

The budgetary pressure of caring for the growing elderly population in hospitals and nursing homes has promoted interest in other less costly arrangements. Hospice and home health programs are perceived to be cost effective and are heavily funded at the federal level.

Hospice care is intended for the terminally ill. Most hospice patients receive care in their own homes, but the use of special facilities is becoming more prevalent. In a hospice, an interdisciplinary team of health professionals provides individualized care that emphasizes patients' physical and emotional comfort (i.e., palliative as opposed to curative care), as well as support for family members. Hospices strive for improved quality of life in a patient's final days and death with dignity.

Medicare introduced hospice benefits in 1983, but higher reimbursement rates in 1989 accelerated growth in the number of hospices. Higher reimbursement rates increase significantly the number of Medicare-certified providers, improving access for Medicare beneficiaries. In 2012, about 3,800 Medicare-certified hospices served about 1.3 million patients. With evidence that hospice programs offer savings, many private insurers also have added coverage for hospice care.

Home health care, the larger and more important program, provides care to patients with acute and long-term needs, including those with disabilities, those recuperating from a hospital stay, and even the terminally ill. The home care benefit was included in the initial Medicare legislation of 1965 that was extended in 1973 to certain disabled persons under 65. By 1996, about 2.4 million home health care patients were on the rolls of agencies at any time, and their numbers were growing rapidly. With a doubling of patients in just four years between 1992 and 1996, home health care became one of the fastest-growing components of total spending. Between 1992 and 1996, Medicare spending for home health care grew from $7.7 to $18.1 billion. Medicare spending, however, fell to $7.6 billion by 1999 following the Balanced Budget Act of 1997. As with nursing home care, the act mandated a shift from a cost-based to a prospective-based system of reimbursement, and installed an interim system of payment limits over the intervening period. By 2000, there were only 1.4 million home health care patients at any time and national spending on home health care stabilized at $32 billion. Since then, spending has again climbed rapidly with over 12,000 home health agencies serving nearly 5 million clients. Total home health care spending reached $80 billion in 2013 with Medicare and Medicaid accounting for 80 percent.

The rationale for public funding for home health care rests on the premise that it is much less expensive than either hospital or nursing home care. Even though a home health visit is unquestionably far less costly than a day spent in an institution, the effect on total health spending is not entirely clear. The principal issue is the extent to which home health substitutes for "unpaid" care by family members and other caregivers, or for institutional care.[7]

Policymakers are concerned that more generous public funding for home care will substitute for previously "unpaid" care without significantly increasing the overall care for patients. Yoo and colleagues (2004) have shown that formal and informal LTC are close substitutes. From data for 15 OECD countries, they estimated that the availability of a spouse caregiver, measured by the male–female ratio, reduces LTC spending by $29,000 (in $1995), a figure that exceeded the annual Medicaid payment to an intermediate care facility.

The substitutability of informal care with LTC is confirmed by Van Houtven and Norton (2004), who introduce a utility-maximizing framework to analyze the informal care decisions by children and their elderly parents. Children select the optimal amounts of consumption, leisure and informal care, subject to their budget constraints. The utility function includes the parent's health status. The utility functions maximized by parents include consumption, formal medical care, and the informal care available from their children. The simultaneous maximization process predicts that informal care could be either a substitute or complement to formal care, and that the substitute or complement effects may vary with the type of formal care consumed (e.g., substitute for LTC and complement to inpatient/outpatient hospital care). The results of their empirical estimation indicate that informal care provided by children is a net substitute for both LTC and hospital care and physician visits. It is a complement to outpatient surgery.

Does it matter whether the caregiver is a son, daughter, or someone else; whether the adult child is married; or whether the recipient is married? Subsequent work by the same authors (Van Houtven and Norton, 2008) focused on these and other differential effects. Informal care by adult children is a net substitute for skilled nursing home and inpatient care. Children are less effective for married than for single recipients and children of single elderly are more effective than other caregivers. However, the gender of the adult child caring for a parent does not matter.

The policy and cost implications of the significant substitutability of informal care for LTC are substantial. Social changes that reduce children's abilities or commitments to care for their parents would further strain private and public budgets. Conversely, well-designed programs that provide incentives for children or other family members, especially those who are not in the labor force, to care for parents could lead to substantial government savings.[8]

Conclusions

Following a discussion of the history and organization of hospitals and hospital costs, we examined two economic and policy issues—the medical arms race (MAR) hypothesis and cost shifting. Careful analysis indicates that both issues are far more complex than they first appear. Common perceptions of a wasteful MAR and complete cost shifting do not accurately represent how hospital markets function. We also examined the determinants of consolidation and exit in the hospital industry, including the effects of the ACA, and some of the consequences of this restructuring. There are significant deficiencies in the quality of hospital care, and improving quality is one of this nation's most important goals.

In the long-term care sector, we focused on nursing homes, emphasizing three issues: quality, especially for Medicaid patients; shortages; and financing nursing home care. We found that economic theory and empirical evidence can provide useful and sometimes surprising results. For example, no clear relationship exists between costs and quality. It is also possible to have a persistent shortage of nursing home beds without any mechanism, such as price, that would alleviate the shortage.

Finally, we examined potentially less costly alternatives to nursing homes and the role of informal care. Home health care represents one of the fastest-growing components of health care spending. Informal care and LTC are close substitutes, so policies that encourage informal care could substantially reduce public spending for LTC.

Summary

1 The modern hospital evolved at the turn of the twentieth century following the invention of the X-ray and significant advances in antisepsis, anesthesia, and the biological sciences.
2 Hospital spending has grown rapidly in recent decades as a result of the growth of private and public insurance and other factors. It accounts for about 32 percent of national health expenditures.
3 The hospital industry has experienced rapid change, including reductions in the number of hospital beds and inpatient utilization and significant growth of outpatient services. Hospitals are facing competitive pressure to restructure through mergers, participation in hospital networks, and other partnerships.
4 Hospitals are licensed and subject to a wide range of state and federal regulation.
5 Many analysts believe that the hospital industry is in a medical arms race resulting in unnecessary duplication of expensive technology. The limited empirical evidence does not support this view.
6 Intuitive reasoning suggests that the costs of discounts or uncompensated care to some patient groups must be passed on to other paying groups. More formal analysis leads to a richer set of results including situations where costs cannot be shifted. The empirical literature indicates that cost shifting is far from complete and the welfare loss associated with any shifting is relatively small.
7 Increased market power, rather than improved efficiency, has been the principal driving force for hospital consolidations. Less efficient and less profitable hospitals are more likely to exit. The ACA has created financial incentives that are fueling further consolidation.
8 Improvements in the quality of hospital care, and the role of publicly available performance indicators, have emerged as major policy and research themes. To improve quality of care, the ACA levies penalties on hospitals that do not meet various quality standards.
9 Patients are concerned about and react to hospital quality differences.
10 The nursing home population has grown dramatically since the introduction of Medicare and Medicaid. Medicare and Medicaid pay 52 percent of all nursing home costs.
11 Nursing home quality has been examined through structure, process, and outcome indicators. Surprisingly, no conclusive evidence relates cost to quality or supports the view that nursing homes with higher proportions of Medicaid patients produce lower-quality care.
12 A familiar theme in the nursing home literature is one of persistent excess demand by Medicaid patients. A model of chronic excess demand is plausible, although recent evidence indicates that excess demand is not a universal phenomenon.
13 The nursing home population has leveled off, but financing nursing home and long-term care remains a great social challenge. Medicaid has tightened its eligibility thresholds for nursing home care.

14 Home health care and other alternatives are growing rapidly in number of patients and costs. They can be cost-effective alternatives to hospital and nursing home care.
15 Informal care and long-term care are close substitutes. By encouraging or discouraging informal care, policies have significant budgetary implications for Medicare and Medicaid.

Discussion Questions

1 Explain why it is often claimed that hospitals compete for doctors rather than patients. What are some of the implications of this phenomenon, assuming that it is true?
2 Even nonprofit hospitals must earn a "profit." Evaluate this statement.
3 What is the medical arms race (MAR) hypothesis? What features of hospital markets make the presence of an unproductive MAR possible?
4 Suppose that the Medicare rate of hospital reimbursement is reduced. Explain why the costs may not be shifted to other patients in the short run.
5 Explain why only about 5 percent of adults buy long-term care coverage.
6 Hospital costs have grown following the growth of private and public insurance. Describe other factors that could account for some of the growth.
7 The headline of an August 21, 2005 article in the *New York Times* was "It's the Simple Things, but Some Hospitals Don't Do Them." Use the "Quality of Care" section to discuss and explain why hospitals may fail to provide some simple and effective life-saving procedures.
8 The article in Box 14.2 describes similarities between the hospital and airlines industries. What are some significant differences that may prevent or minimize some of the outcomes for hospitals that are suggested by that article?
9 Explain how excess demand for nursing home beds may persist over long periods. How can the hypothesis be tested?
10 Nonprofits are dominant in the hospital industry, while for-profits dominate the nursing home industry. Develop some possible explanations for this difference.
11 Informal care provided by children and other family members are good substitutes for LTC for parents. Describe some potential social and demographic changes that may reduce the availability of such informal care. Develop policies that may help take advantage of the substitutability to delay entry of the elderly into LTC facilities.

Exercises

1 What is a dominant strategy in game theory? Using the payoff matrix shown in Box 14.1, replace the entry in the first row and second column with (125, –50). Does A have a dominant strategy? Does B? What is the solution to this game?
2 Assume that there are three groups of hospital patients (instead of two as shown in Figure 14.1): private, Medicare, and Medicaid (which has a lower fixed rate of reimbursement than Medicare). Explain how a hospital would select patients in order to maximize profits.
3 Suppose that Medicaid's hospital reimbursement rates do not cover the variable costs of patient care. Will a profit-maximizing hospital accept Medicaid patients? If not, under what circumstances will the hospital accept such patients?

4 Assume that a patient's health can be improved by home care or institutional care. Use isoquants and isocost curves to determine the condition for efficient utilization of these two forms of care. Under what circumstances will an insurance program promote or fail to promote efficiency?
5 Consider Figure 14.1. Suppose that a hospital has the ability to be a perfect price discriminator in the private market, that is, it can charge every private patient the maximum that the patient is willing to pay. Explain how this will affect the number of private patients the hospital will take and the impact on the hospital's profits.

Notes

1 More extensive discussions of many of the topics in this section are found in Starr (1982), Temin (1988), and Raffel, Raffel, and Barsukiewicz (2002).
2 American Hospital Association, *Chartbook* Table 4–4 (aha.org/research/reports/tw/chartbook/index.shtml: accessed January 26, 2016).
3 With capacity limitations, it will first raise the price to eliminate those private patients whose marginal revenue is below the Medicare rate.
4 Cost shifting could arise if the hospital was not previously maximizing profit and was accepting "unprofitable" patients whose marginal revenue fell short of marginal cost. The lower Medicare rate would encourage the hospital to reduce the number of these patients by raising the private rate. Santerre (2005) describes the welfare loss that would arise. His estimates for 1992 indicate a maximum welfare loss of just 0.84 percent of private hospital expenditures.
5 The rate that would eliminate the Medicaid excess demand is found by sliding the segment $A'C'$ further up the marginal revenue curve until the quantities such as $B'C'$ are eliminated. Clearly, the rate must be above R_2.
6 CON regulations require health care providers to obtain approval from state planning agencies for capital expenditures exceeding a threshold level such as $500,000. CON was federally mandated until 1987 and then left to the states.
7 An extensive literature is available on the impact of home care on hospital utilization. Hughes's (1997) meta-analysis found that home care unambiguously reduces hospital days for the terminally ill. For others, the effect on hospital days is negative though not overwhelmingly large. See also Forder (2009) and Bonsang (2010) for analyses of various substitution effects in European countries.
8 Economic analysis also suggests that policy needs to recognize the adverse effects that informal care giving may have on family and friend caregivers. Using self-rated life satisfaction to measure well-being, Berg, Fiebig, and Hall (2014) found that the negative effects can be substantial.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 15

The Physician's Practice

In this chapter

- Physician Agency and Supplier-Induced Demand (SID)
- Small Area Variations (SAV)
- Issues that Affect Both SID and SAV
- Conclusions

In the United States, private practice physicians generally operate in for-profit firms, so it might seem a simple task for economists to model their market behavior. In this chapter, however, we investigate two physician practice characteristics that distinguish them from other for-profit firms. First, physicians are agents for their patients, agents who enjoy a vast, asymmetric information advantage over their clients. Will some physicians abuse this advantage for private gain? Health economists have taken this question seriously for many years, and we examine the progress of theory and evidence. Second, physician utilization rates often differ substantially across small geographic areas. In this chapter we address the issue from a supply and demand analysis described by Skinner (2012), and we update the empirical evidence.

Physician Agency and Supplier-Induced Demand (SID)

On becoming ill, consumers hire health care professionals to serve as *agents*. Agency is common in fields like auto mechanics, law, and medicine where professionals have a great deal more knowledge of the subject at hand than the consumer. When parties have unequal knowledge, we refer to the problem as *asymmetric information*, as we discussed in Chapter 10. In medicine, we identify the physician as the *agent*, and the patient as the *principal*.

The policy concern is that out of self-interest physicians may violate their roles as agents. Economists have defined the "perfect agent" as one who makes those choices and recommendations on behalf of the patients that the patients themselves would have made if they had the same information. We will describe physicians who knowingly induce their patients to consume other than this optimal amount of care as being in violation of agency. There are matters of degree, certainly. We would characterize the physicians who induce their patients to have an unneeded and risky heart surgery as outright fraudulent. However, if a physician recommended an unneeded follow-up visit, while technically a violation of agency, it would not warrant public oversight.

It is logically possible to observe supplier inducement that entails no violation of agency. For example, a physician may encourage a patient to exercise more or undergo diagnostic screening more frequently. Inducing more care does no one harm if it encourages a move toward the patient optimum. Inducements such as better office décor or more personal care represent complements to the physician output. Such complements provide utility and are part of the package among which a well-informed patient, hypothetically, would evaluate his optimum. More personal care, for example, may even increase the probability of good health; in this case the complement to care in question is an increase in the quality of care.

Modeling Supplier-Induced Demand

Health economists have modeled supplier-induced demand (SID) for at least two reasons. First, one wishes to understand the motivations of physicians, how their incentives affect their practice. Second, models help to understand the data we observe. Furthermore, scholars once debated over the question of whether evidence of "inducement" was consistent with the neoclassical model of markets. To begin, let us acknowledge that the asymmetric information advantages give the physicians the power to misuse their agency relationship with the patient to personal advantage.

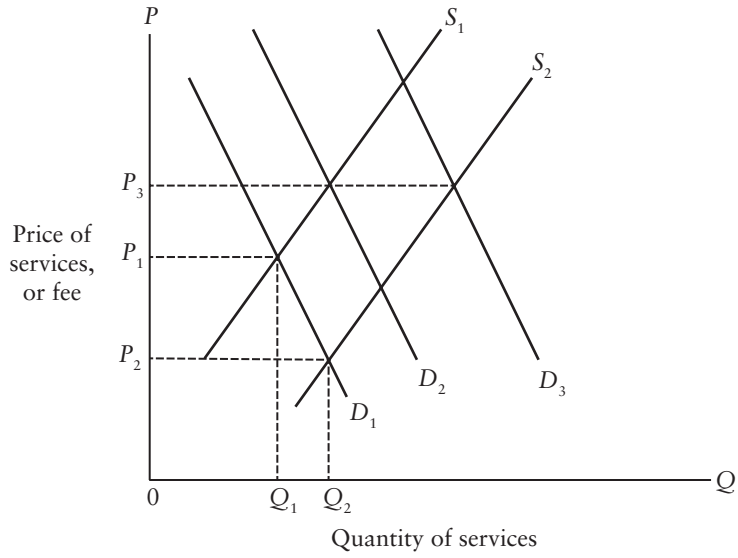
The Supply and Demand Model

If all physicians and patients in the market behaved like a standard supply and demand model, an increase in the supply of physicians would lead to an increased aggregate quantity of care as shown in Figure 15.1. Note, however, that one need not assume SID to predict aggregate demand increases in response to increased competition. The simple market supply and demand model also predicts this. From Figure 15.1, we note that an increase in supply from S_1 to S_2 implies an increased quantity consumed from Q_1 to Q_2, and a decreased price (or fee) from P_1 to P_2.

Uwe Reinhardt proposed a "fee test" to address this confusion, claiming that a rise in physician fees to a level higher than before the supply increase could be caused only by inducement. This is an interpretation of the shift in demand D_3, which is sufficient to cause fees to rise from P_1 to P_3. However, Feldman and Sloan (1988) showed that a model incorporating quality as a demand determinant can explain the same phenomenon. If physicians respond to competition by increasing their quality, and if that higher quality earns a higher price, then SID is again not needed as an explanation.

Do Physicians Respond to Profit Incentives?

To prove that physicians practice SID, we would need to show that they respond to profit expectations. This is a necessary condition but not a sufficient one. There is ample evidence that physicians do respond to profit incentives. One sort of evidence describes physician output changes when facing clearly different reimbursement methods. For example, when physicians are paid per service provided, they provide more services than when they are



Note: The supply and demand model shows quantity increases resulting from a supply increase. The Reinhardt fee test argues that SID is identified when demand increases sufficiently to cause the price (fee) to rise.

Figure 15.1 The Supply and Demand Model of SID

given *capitation*, a fixed total payment (Nassiri and Rochaix, 2006). This makes sense; under fee-for-service, additional treatment means more revenue to cover the increased costs. Under capitation, additional treatment simply means increased costs.

Studies also suggest that physicians respond to income pressures on their practice by striving to increase their incomes. In Norway, patients must register with a physician, creating a patient list for each physician. Physicians with short lists tend to grow their lists more than average during the next five years to make up for reduced income (Iversen, 2004).

Analysts have found that OB/GYN physicians will recommend caesarean section (C-section), a more lucrative (surgical) treatment, when their practice incomes are threatened by competition (Gruber and Owings, 1996). Others find that more profitable settings for surgery tend to be used more (Plotzke and Courtemanche, 2011). One study finds that patients whose physician receives reimbursement under a capitation system may get fewer services and thus perhaps lower quality care (Quast, Sappington, and Shenkman, 2008). There is evidence that physicians have personal goals for income and adjust their practice prices and qualities when these goals exceed their current income (Rizzo and Zeckhauser, 2003, 2007). There is also evidence that when government provides incentives for physicians to choose more cost-saving methods, physicians will tend to save costs (Ho and Pakes, 2011).

These studies support the idea that physicians respond to financial incentives. They also raise a deeper policy question: Does induced demand lower patient well-being, as is suggested in SID theory? The model, Figure 15.2, shows the issues when physicians profit from inducement at the same time that inducement gives the physician disutility.

Consider the trade-off between net income π and inducement I. The model proposes that physicians dislike inducing patient demand, viewing such activity as "less than professional." With each unit of induced patient care, the physician experiences a decline in utility that must be offset by the extra utility gained from the extra income that inducement brings.

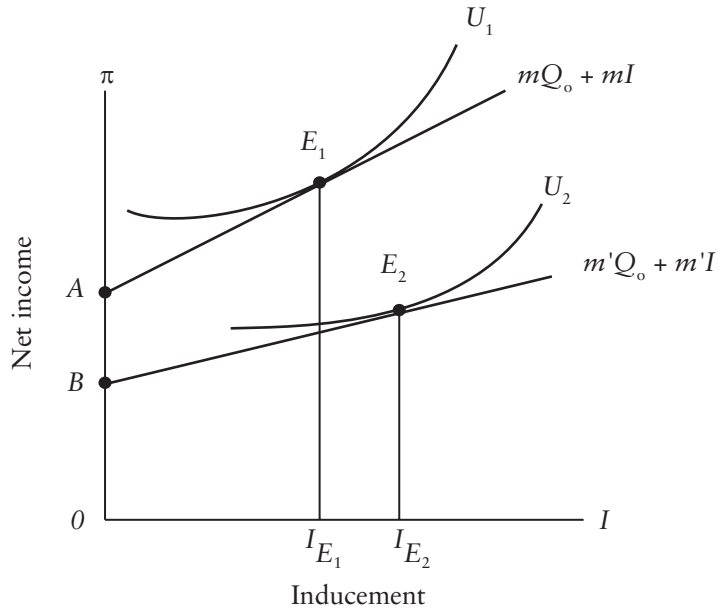


Figure 15.2 Physician's Response to Reduced Rate of Profit

Figure 15.2 represents the process of choosing levels of net income π and inducement I. The indifference curves represent the physician's preferences. The indifference curves slope upward because one of the two "goods," I, is really a "bad." To remain on the same indifference curve, the physician must gain additional net income to offset the disutility of engaging in a higher level of I. As is the usual case, however, higher curves are preferred.

In Figure 15.2, net income π with zero inducement is at point A. This point represents mQ_o, where m is the assumed profit rate (related to, but not necessarily identical to, the physician's wage) from each unit of patient care, and Q_o is the amount of patient care with zero inducement. Net income π increases by rate m with each unit of inducement I along the income line $mQ_o + mI$. The physician's initial equilibrium is determined by the tangency of the net income line and the indifference curve, shown at E_1, where the physician induces OI_{E_1} of extra patient care.

With a lower profit rate $m' < m$, the net income line becomes flatter and lower, or line $m'Q_o + m'I$. The physician now chooses equilibrium point E_2. In this case, it results in OI_{E_2}, a higher level of inducement than OI_{E_1}, even though the doctor had a higher profit at E_1.

The model describes physicians who can induce demand but dislike doing so. To what extent would physicians go to overcome this disutility for financial gain?

The Target Income Hypothesis

Economists who first proposed the SID criticism of physician behavior formulated the "target income hypothesis" (Evans, 1974). This argues that physicians have desired incomes that they strive to achieve or to restore whenever actual income falls below the targets. This target income model is a relatively extreme one.

Target income behavior suggests that for the physicians in question, income becomes not merely the main thing but the only thing. This extreme focus on an income target, as well as the inflexibility of the target, were features that caused many health economists to question the idea.

An empirical question: Would physicians adjust their price and quantity decisions if their target income changed? Suppose we asked physicians their desired income. Do physicians who express levels much higher than their current income choose different and more profitable price and quantity combinations than others? Evidence (Rizzo and Blumenthal, 1996; Rizzo and Zeckhauser, 2003) finds that they do. Furthermore, physicians appear to change their price and quantity choices when they make gains in income from sources unrelated to their practices (for example, stocks and bonds). For one physician's perspective, see Box 15.1.

BOX 15.1

SID and Target Income: A Physician's Perspective

The models of SID and target income necessarily abstract from the care process rendered by skilled professionals who are seeking to provide the most appropriate treatment for their patient. This personal perspective comes from an assignment in a health economics course at a local university. The writer is a physician who is director of pulmonary critical care at a large hospital.

I would, however, agree that physician behavior does and will change in response to income fluctuation but I would lend some insight into the current thought process. Recall, always, that the physician does not determine charges or the criteria for reimbursement for inpatient or outpatient services. The physician does not "approve" of the levels of service for an inpatient or the criteria to an MRI, CT, or PET scan. Insurance companies and Medicare determine all these rules. By and large, most physicians attempt to provide the best available care, but now being cognizant of the myriad of work needed to support a claim.

Let me give an example of how this works in my private practice. I am asked to see a patient in consultation. The patient may have been post-op with respiratory failure. Without regard to patient insurance, I provide care and dutifully document my activities with the patient. After the patient recovers, I post a billing, let's say to Medicare. There was an initial consult and 10 days of care in the ICU and step-down unit. Based on my provider ID number (PIN) my claim is categorized. Routinely, I get 10–15 percent rejection due perhaps to incomplete data transmission or another doctor who also billed a similar "category" code. I then must use my staff to copy the progress notes in medical record and provide these documents to the payer. The delay is now perhaps 90–120 days.

Have I put my income interest above the patient's best interest and welfare? A resounding No!! However, what physicians are actively pursuing is the details and methodology to code and document properly so as to receive payment for what has been actually done . . . I favor full prosecution of fraudulent behavior, but there is a distinct philosophical difference between billing what was actually performed and understanding the nuances of the billing coding, reimbursement world as contrasted to a doctor that would purposely "up-code" billing for services not rendered and deliberately mislead a patient. Institutions are also working aggressively to accurately document patient care and co-morbidity to increase (read legally!) reimbursement and document a higher case-mix index. Continued efforts by payers to monitor services, prescription patterns, and key benchmark care points will not only help rein in any outliers but also improve patient care and safety by eliminating wasteful care.

The McGuire and Pauly Model

This model captures most of the observed physician behaviors as special cases of a utility-maximizing physician decision maker. We can conceive of a profit-maximizing firm as a utility maximizer who emphasizes profits *far beyond other goals*. In the McGuire-Pauly (1991) model, the physician gets utility from (1) net income and (2) leisure, and disutility from (3) inducement, the physician's own efforts to induce patients to buy more care than appears medically necessary. This last factor introduces the supplier-induced demand (SID) controversy into the model. As we will see, the question regarding SID is whether physicians use their knowledge advantage to abuse their agency role for monetary gain.

Let the physician's utility function be:

$$U = U(\pi, L, I) \tag{15.1}$$

where π is the net income from the practice, L is the physician's leisure time, and I is the degree of inducement. The physician can choose any amount of labor and inducement effort consistent with the profit level implied by these choices.

With three variables to consider, the physician must consider the willingness to trade-off between three pairs of goals:

Net income π and Leisure L;
Leisure L and Inducement I;
Net income π and Inducement I.

The overall decision involves all three variables, but a good understanding of the model focuses on two pairs. First, consider the trade-off between income and leisure. In panel A, Figure 15.3 assumes that the work hours return a constant revenue, w, for each hour worked. As depicted, a physician's (after-tax) "wage," w, determines the slope of this labor–leisure trade-off. Starting from the maximum possible leisure (the horizontal intercept in the graph), each hour worked represents one hour less leisure. Corresponding to this, income rises by w, the net payment per hour worked. It follows that higher wage levels $w_2 > w_1$, and then $w_3 > w_2$, will result in steeper rising income lines. The physician chooses the optimal points on each income line; the indifference curves in Figure 15.3 illustrate this process.

Note that the path of optimal points—from A to B to C—at first heads to the northwest and then bends back and upward to the right. This pattern is recast in panel B, where wage levels (w_1, w_2, and w_3) and labor are on the axes, in this case forming the "backward-bending labor supply curve." As conventionally explained, in the region from A' to B', the physician is primarily motivated by higher wages to substitute labor for leisure; but, in the region from B' to C', the physician's income effect dominates this substitution effect. The physician becomes rich enough to wish to spend more time enjoying the income.

These graphs show that the physician's income and especially the income effect play critical roles in determining how much he or she is willing to work. In panel B, suppose a physician is

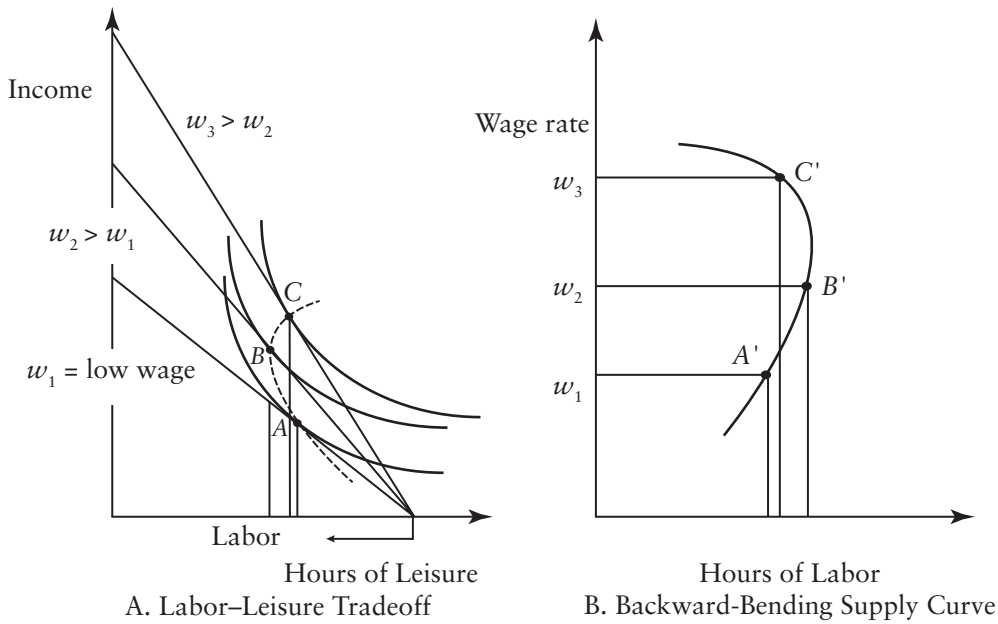


Figure 15.3 Supply of Physician Labor

at point C', and then the wage rate falls. The physician shown would move along the curves toward point B'. This implies a wish to compensate for lost income by working more.

We can merge the various SID versions using the model. Figure 15.4 compares target income behavior in panel A with profit-maximizing behavior in panel B. In panel A, once competition forces the profit rate m to a lower level, m', the new equilibrium is tangency point E_2.

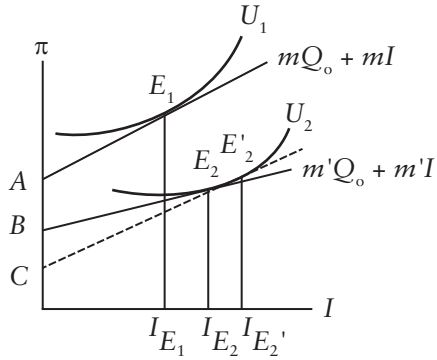
The broken line identifies the physician's changed inducement in a different way. It hypothetically removes income from the physician who is at E_1 until the physician attains the equivalent utility to E_2. The resulting change in inducement caused is the *income effect* and it is measured in this case as $I_{E_1}I_{E_2}'$.

Panel B shows the contrasting profit-maximizing behavior. Notice that the income effect in this case is zero; we see this in panel B because the equilibrium inducement does not change when income is removed. Similar to the previous analysis, the broken line shows this tangency at E_2', which equals the inducement level at E_1. Why is a zero income effect like the profit maximizer? The profit maximizers gain utility only by the net dollars brought in, and the decisions made are unaffected by other matters, such as their incomes. As a result, the profit maximizers' income effect will always be zero. Notice in addition that the physicians in panel B would reduce inducement when faced with greater competition and a lower m. The new equilibrium is at E_2. In this case, it results in OI_{E_2}, a lower level of inducement than OI_{E_1}.

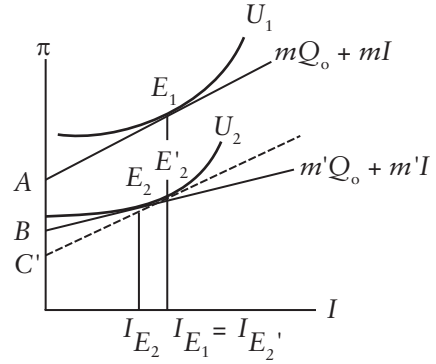
The McGuire-Pauly model explains that the size of the income effect is critical to understanding and identifying SID behavior. A lower profit rate, m, has two offsetting effects on inducement:

> *Substitution effect*: If inducement is less profitable (smaller m), providers would do less inducement, that is, substitute away from it.
> *Income effect*: Decreased income would make inducement more desirable.

For inducement to increase, a positive income effect on inducement must be large enough to overcome the negative substitution effect on inducement. This fact has led analysts to search for evidence of a substantial positive income effect.



A. Physician responds to lowered profit rate m' by increasing influence on patient demand.



B. When there is no income effect, as in profit-maximizing behavior, then the model predicts a decline in supplier influence will follow a decrease in the profit rate.

Figure 15.4 The McGuire-Pauly SID Models

What Do the Data Say about Supplier-Induced Demand?

Two criticisms were raised about much of the early SID work. First, many of those studies could not distinguish between the SID model and the conventional supply and demand model. Second, many estimates of the SID effect proved to be statistically flawed, meaning that the econometric coefficient thought to be evidence of SID could not be disentangled from other coefficients. In Box 15.2 Dranove and Wehner provide an example of such flaws by looking at the "supplier-induced pregnancies."

Re-examining the competitive model in Figure 15.1, we see that with a sufficient degree of SID the physician's fee level can rise in response to greater competition. Recall the Reinhardt fee test for SID, which detects the presence of SID via its effect on physician fees. The premise of this approach is that price cannot rise in response to increased competition unless there is SID. However, Figure 15.1 applies precisely only to the perfectly competitive market structure. Many prefer to describe physician markets as monopolistically competitive with the firm having a downward-sloping demand curve. McGuire (2000) showed that the implications of availability on fees in that case are not so clear. In addition, if physicians can adjust their quality in response to increased competition, then higher fees could result even when there is no SID (Feldman and Sloan, 1988).

BOX 15.2

Supplier-Induced Pregnancies

Dranove and Wehner (1994) challenged the accuracy of standard statistical methods used by health economists to test for SID in an unusual and revealing way. They deliberately assumed—counter to all logic—that obstetricians can influence demand for the delivery of babies. They further "assumed" that an increase in obstetricians per capita will lower the expected incomes of these obstetricians and impel them to use their superior knowledge vis-à-vis the patient to induce demand for childbirths. The two then applied a typical SID approach to investigate.

First they showed that availability of obstetricians and gynecologists per capita and childbirths per capita were positively correlated, an interesting result but one consistent with the ordinary long-run operations of supply and demand. They then measured the "pure SID" effect. Their estimated SID elasticity for obstetricians per capita was 8 percent and significant. Using this figure in their further calculations they found that a one standard deviation influx of obstetricians will induce an additional 7 percent in childbirths per capita. Can we conclude that the obstetrician newcomers were, let's say, socially very active? Hardly.

Gruber and Owings (1996) also applied the fact that pregnancies are determined by parents and not by the doctors. Reasoning that the fall in fertility among U.S. women was unrelated to their model of obstetrical care, they measured the change in caesarean deliveries (C-sections) relative to vaginal deliveries, C-sections being more lucrative for obstetricians. They concluded that the drop in fertility and consequent effects on physician income led to an increase in C-sections. Physicians were able to recoup about 10 percent of their income drop by encouraging the C-sections.

Addressing similar questions, Norwegian researchers investigated the physicians' responses when they have shorter than average lists of patients. These physicians tend to grow their patient lists faster than average in subsequent periods (Iversen, 2004; Grytten and Sorensen, 2007, 2008). However, the physicians with short lists did not increase service production per consultation.

Nguyen and Derrick (1997) found that Medicare fee reductions led to increased volumes of care for the physicians in their sample who incurred the greatest fee reduction. Yip (1998) likewise found that Medicare fee reductions (in New York state) led to increased volumes.

One study reports on patient utility. If we rely on patient self-reports, we get another provocative contrast. Carlsen and Grytten (2000) found that patients in Norway enjoyed *greater* consumer satisfaction after increases in the market area's availability of physicians. Plausibly, each physician has a somewhat smaller workload and makes up for this in part by spending more time with each patient. The two authors reason that if reported consumer satisfaction adequately reflects patient utility, then SID may not matter to patient welfare.

This section asked "What do the data say about SID?" It is clear from these studies that reductions in net income led physicians to induce demand in many cases. However, none of these cases necessarily demonstrates physicians violating their agency roles. For example, improving quality to attract more patients would not necessarily violate his trust with current patients.

A Marketplace Approach

Critics of SID have often complained that evidence of inducement is mistaken for other phenomena. For example, we know that pregnancies are not induced by an influx of OB/GYN physicians. A related criticism is that the effect of an influx of physicians is often not statistically identified, meaning that the estimating model could not separate the SID effect from other influences. Were there any valid evidences of SID? Yes. That too is clear. But now many health economists explain SID as merely one of many forms of marketing, which virtually all firms use. The theme of this section is that while SID is harmful to social welfare, many forms of inducement may be beneficial.

A model proposed by Feldman and Sloan (1988) showed that when quality change was possible, physician firms might respond to increased competition by choosing a higher quality level. This explains a higher price without resorting to the theory of SID. What could that higher quality be? The firm could employ more Board Certified physicians, purchase better diagnostic equipment, encourage nurses and other staff members to treat patients better, or rent a nicer office building.

Others noted a similarity of physician influence with commercial advertising. In 1985, Uwe Reinhardt developed a model of health care advertising that would in given circumstances both lead to increased demand for the physician firm as well as a higher equilibrium price in response to competition. Miron Stano (1987) extended the advertising idea to conclude that the price could rise and quantity could change, but each depended on the market structure.

A drawback of the advertising analogy is that most physicians do not seem to like advertising. In fact, some theorists have described advertising as a signal of low quality. This negative view echoes in part the marketing theory that advertising can be of two types: informational or emotive. Perhaps products and services more substantive than soft drinks, for example physician's services, require more informational content for advertising to be effective.

Suppose instead that we characterize advertising to incorporate all the demand-inducing mechanisms frequently described by marketing theorists. These include adjustments in price,

changes in the product or service, or product promotion. This latter concept includes not just advertising but also the kinds of physician recommendations that have been the focus of SID literature.

Are these common marketing practices the same as SID? Health economists distinguish between inducements that primarily benefit the patient ("quit smoking" or "take these vitamins and see me in three weeks") and those that primarily benefit the physician ("you need knee surgery" but it is not medically indicated). We take this last as the definition of SID. If doctors caught a colleague doing this, they would depict the behavior as fraud. Yet SID is often not so clear, and it is always difficult for economists to detect with econometric tools. Many influences inspire patient demand such as improved quality of care, clinic amenities, or accessible location.

The confusion is illustrated in the extreme by a Norwegian study. A physician's response to a decline in demand was to increase the intensity of service, charge more, and restore some of his income; however, a survey of his patients found them to be pleased with the change (Iversen 2004). Does SID as critically defined actually exist? It clearly does. Recall for example (Gruber and Owings, 1996) the study showing OB/GYN physicians' response to increased competition by inducing greater use of the C-section, a more lucrative procedure. Is all physician demand inducement harmful to the patient? It seems not. Inducement may encourage a patient to move in the direction of the optimal, or it may work its inducement by improving the quality of service.

Conclusion on SID

We have looked at physician inducement first in the prominent models. Recall that in those approaches the physician can induce demand but experiences disutility by doing so. This disutility is the cost of inducement, an essential element of the model; if it had no cost the physician would always induce to the max—which makes no sense. Notice that in the marketplace the real cost of improving the quality of services performs this same function. Finally, here inducement is not automatically equated with harm to the patient.

To conclude, it is difficult to identify the prevalence of SID econometrically in a market. Are some physicians fraudulent? Surely there are bad people in every profession, but after years of study, health economists have found SID to be less of a problem than many once thought.

Small Area Variations (SAV)

Modern small area variation, or SAV research, stems from the pioneering work of John Wennberg and colleagues, who studied New England hospital markets. Later studies corroborated their evidence of wide variations across small service markets. Why did women in one New England town undergo hysterectomies at more than twice the rate of another apparently similar New England town? Understandably the variations proved worrisome, and many researchers focused on discovering their sources.

The favored measure of small area variations has been the coefficient of variation (CV) and the systematic component of variation (SCV). The coefficient of variation $CV = sd/mean$ divides the standard deviation, *sd*, of the observed medical use rate by the *mean* of the same measure. Dividing *sd* by the *mean* adjusts for the size of the rate being studied. For example, the unadjusted rate for treatment of Guillain-Barré syndrome (a relatively rare disorder,

afflicting only about one person in 100,000) would appear to be much smaller than the rate for treatment of the common flu. Division by the mean rate of each serves to put these on a common basis.

The SCV is a twist on this theme. Here, the researcher first removes from the observed treatment rate that portion of its variation that we can attribute to the random nature of the disease. The remaining portion then becomes a better measure of the effects of systematic factors, such as physician practice styles and supply and demand factors.

Contributions to These Variations

Table 15.1 excerpts data from a study by Wennberg (1990) featuring the CV statistic. Phelps (1997) characterizes a CV in the range 0.00 to 0.10 as low variation, 0.10 to 0.20 as moderate, and over 0.20 as high. Of the ten ailments reported, most of these procedures would be considered highly variable based on the size of their CV statistic. The *extremal* ratio, also reported in the table, while not as reliable a measure of variation, is easier to visualize. It is the ratio of the largest utilization rate observed across the areas to the smallest. Thus, the rate of mastoidectomy across the study areas varies by a factor of over four.

Researchers have investigated these variations. Much of the SAV work focuses on the contribution of socioeconomic characteristics of the population and the role of the availability of supplies of hospital and physician services (see Alexander et al., 1999). The studies together reached two conclusions: (1) supply variables are important and demand characteristics play a somewhat lesser role, though both are statistically and materially significant; and (2) such variables do not seem to suffice, as much variation is unexplained (Folland and Stano, 1990).

Wennberg (1984) argued that much of the observed variation relates closely to the degree of physician uncertainty with respect to diagnosis and treatment. When there is little consensus about the effect and value of a medical procedure, a wide range of physician treatment choices lies within the bounds of accepted practice. The physician's habits of treatment

Table 15.1 Variations by Medical Procedure Category

Surgical Procedure	*No. of Cases*	*Coefficient of Variation*	*Extremal Ratio*
Colectomy	3,190	.116	1.47
Open heart surgery	1,439	.232	2.29
Appendectomy	5,381	.305	2.86
Thyroidectomy	949	.342	3.35
Total hip replacement	1,717	.353	2.99
Diaphragmatic hernia	2,178	.369	3.45
Coronary bypass surgery	3,744	.383	3.62
Mastoidectomy	569	.461	4.03
Spinal fusion w/wo disc excision	1,234	.520	5.20
Total knee replacement	998	.525	7.42

Source: Based on information from Wennberg (1990).

choices, beliefs about efficacies, and patterns of practice are said to determine her practice style. Phelps (2000) speculated that physicians differ in their patterns of practice because of the imperfect diffusion of information on medical technologies. The medically optimal production function may not be widely known. Some physicians may be optimistic about the good effect of a given procedure, yet others pessimistic about it.

Furthermore, as medical science uncovers better information about the true production function, efforts to disseminate this information will help reduce unnecessary surgery. Promoting medical practice based on scientific findings is called *evidence-based medicine*. The ACA, which greatly expanded health insurance coverage in the United States urges empirically based medicine as an element of cost control. Reliable quantitative estimates are hard to come by, but some experts claim that the portion of ineffective treatments is high.

Education, Feedback, and Surveillance

Studies show that information programs directed at physicians can alter their behaviors. One early study (Wennberg and Fowler, 1977) found that an informational program significantly affected the tonsillectomy rates in 13 New England areas. Another (Dyck et al., 1977) found that the rate of "unjustified" hysterectomies dropped by two-thirds subsequent to a review program introduced in the Canadian health system. Yet another (Chassin and McCue, 1986) found reductions in unneeded use of pelvimetry following a physician information program.

Epstein and Nicholson (2009) modeled the sources of relevant information. They find that variations between physicians *within* a small area are greater than variations between areas. Contrary to prior hypotheses they find that a physician's location of medical residency has little influence on practice style. Stronger influences are peers within the hospital where she practices as well as peers in the other hospitals in the region. Their study's indicator of OB/GYN practice style was the portion of deliveries conducted by caesarian section. By controlling for patient and physician characteristics, they found that the practice of physician peers was important to the choice of caesarian birth.

Information flows from medical science and from comments by respected peers clearly influence physicians' pattern of practice. Yet, economists point out that supply and demand factors also influence the variations, much like they do for other goods and services (Skinner, 2012).

The Demand Side

First, health status differs substantially across the U.S., as we have seen in Chapter 5, and health status is the main reason why people visit their doctor. However patients often show up for minor or even trivial reasons, making these into social visits. Further, some analysts find that a large share of medical treatments have little or no effect on health status. Thus the association of health status with health care utilization may be weaker than desired.

Second, risky behavior varies across the U.S. For example, alcohol consumption is high in the Upper Midwest, but low in the Southeast. While excessive drinking is clearly risky for individuals, geographic averages can be misleading.

Third, health insurance coverage in the U.S. differs geographically substantially although the Affordable Care Act promises to sharply reduce the percentage of uninsured. Since many variables can affect physician utilization rates, it may be best to approach the problem with multivariate techniques. As analysts add more variables to the analysis, one learns better how much of the observed variation derives from demand factors.

Phelps and Parente (1990) found that standard demand and supply variables typically account for between 40 and 75 percent of the variation in their study of 134 separate diagnostic categories. The authors pointed out that nevertheless these results leave a substantial amount unexplained.

The Supply Side

In the SID section of this chapter, we encountered substantial complexity in tracing the incentives for observed physician behavior. It is no less complex to trace physician influence on small area variations. Phelps (2000) proposed that imperfect information flows limited the local physician's knowledge of the optimal medical production function. Skinner (2012) described the supply side more fully, emphasizing many supply influences on SAV.

1 Physician financial incentives will vary across areas.
2 Capacities of hospital beds and medical equipment may vary.
3 Patient access to care will vary in cost and time.
4 Medical malpractice risk will vary across areas.

Do the small area variations indicate that inadequate care is common? Recall that Phelps (2002) suggested that knowledge of the medical production function might be limited because of imperfect information flows. Skinner's details of supply and demand factors suggest that many of these might not lead to inappropriate care.

These small area variations in medical utilization do not necessarily indicate harms to social welfare. For example Skinner suggested that variations in meat and poultry consumption don't usually evoke such concern. However, many possible causes of SAV do evoke substantial concern, including pockets of poverty, imperfect physician knowledge of the medically optimal treatments, or excess risk-taking in local cultures such as drugs and alcohol.

Issues that Affect Both SID and SAV

Malpractice

On occasion providers make mistakes. The website StateLawyers.com provides a useful description and definition:[1]

> **Medical Malpractice** occurs when a negligent act or omission by a doctor or other medical professional results in damage or harm to a patient.
>
> Negligence by a medical professional can include an error in diagnosis, treatment, or illness management. If such negligence results in injury to a patient, a legal case for medical malpractice can arise against:
>
> - the doctor, if his or her actions deviated from generally accepted standards of practice;
> - the hospital for improper care or inadequate training, such as problems with medications or sanitation;
> - local, state, or federal agencies that operate hospital facilities.

Although legal cases are fraught with nuances, proof of malpractice requires evidence of harm to the patient and negligence by the physician. Most mistakes in patient care go unnoticed by the patient, but a small number of patients sue for malpractice, and some of these suits will be justified.

The problem for physicians is that many judgments are very large and malpractice insurance premiums can be very high for the most suit-prone specialties. For physicians in total, the premiums are less than 10 percent of physician revenues. Health economists take interest in malpractice insurance costs when they threaten to change physician geographic movements in a manner that might reduce social welfare.

Polsky and colleagues (2010) studied hospital discharge data from Pennsylvania, Florida, and New York. They found that higher malpractice premiums tended to increase the rate of exit and reduce the rate of entry of obstetricians. Premium increases averaging 20 percent per year led to a decline in the obstetrician supply of 5.3 percent.

Helland and Showalter (2009), employing national data on physician liability (a measure closely related to malpractice premiums), found that a 1 percent increase in liability leads to a –0.285 percent decline in work hours for physicians. Among physicians 55 or older this elasticity was –1.224.

Robert and Hoch (2007) asked whether increased malpractice premiums led to increased medical expenditure for the consumer. Using Medicare data, they found that greater rates of malpractice law suits increased medical expenditures per enrollee. They presumed that "defensive medicine" (where physicians over-treat to ward off lawsuits) was the cause. In some jurisdictions the extra cost exceeded 25 percent.

Paying for Outcomes

When a customer gets a car repaired, the mechanic usually guarantees the work, and if the outcome is not satisfactory the customer can go back and have the mechanic make things right. Why can't we do the same with physicians?

Dranove and White (1987) have argued that the common physician form of contract stems from both the difficulty of evaluating the health status of the returning patient as well as the fact that unobservable patient behavior is very important to the outcome. How can one tell if the patient's claim of pain is true, and how can one be sure that the patient has followed the treatment regime faithfully, including taking his medicine as prescribed?

BOX 15.3

Clinical Decision Making and Patient Preferences

David Eddy is a pioneer in studying the role of information, medical science, and preferences in clinical decision making. Eddy (1990) breaks the treatment decision into two main components—evaluating the outcome of the alternatives and then comparing and choosing from among the options. The first component is largely a scientific one determined primarily through clinical studies of patient outcomes. However, for many patient conditions, even a fully informed physician will face the problem of inadequate scientific information.

The second component, comparing the alternatives, involves judgment and patient preferences. Eddy emphasizes that patient's preferences should determine the decision (p. 442). A failure to make appropriate decisions can arise from failures in

both stages. Clearly, physicians' inadequacies in education and training, as well as deliberate attempts to misrepresent outcomes (e.g., to engage in SID), will distort information needed at the second step. Similarly, if patients misrepresent their preferences or physicians misinterpret their preferences, or if inadequate information is communicated to patients, the second step will be flawed.

The medical community has come to recognize the role of the patient through the development of information technology to improve the decision-making process. For example, Barry et al. (1995) developed an interactive videodisc that helps patients with prostate cancer choose from among several treatment options. To help with the decision, patients answer questions designed to measure their attitude toward risk and their ability to tolerate discomfort.

Such decision aids have been slow to progress past clinical trials. However, Arburburn et al. (2012) report a Seattle-based Group Health study where patients that received DVDs and educational booklets had 26 percent fewer hip replacement surgeries and 38 percent fewer knee replacements over a six-month period than otherwise similar patients.

Leonard and Zivin (2005) argue that outcome-based contracts will most likely succeed when both patients and practitioners play important complementary roles. This means that they must work together (the provider in diagnosis and the patient in compliance with treatment regimens) in managing or treating diseases, such as diabetes, asthma, or HIV infections. Effort-contingent contracts are likely to be successful when either the patient's or the provider's effort is necessary, but not both; surgery is a good example.

The authors take advantage of a unique opportunity to observe patient choices between fee-for-service and fee-for-outcome of care in Cameroon, investigating patients who chose between the two types of payments. Payment by outcome is rare among physicians in the developed world, but it is common among traditional healers in many African countries. In rural Cameroon, patients may choose mission-based physicians, compensated by physician effort, or traditional healers, compensated by patient outcomes, with both types of providers covering a variety of illnesses. Leonard and Zivin find that when illness requires large amounts of effort by *both* patients and providers, the patients are more likely to seek treatment from traditional healers who are paid based on outcomes. When the disease is not particularly responsive to one of the two types (physician or patient) of effort, patients visit effort-compensated physicians.

Conclusions

Our SID models depict the physician as someone who positively values net income and leisure, and dislikes inducing patient demand. It shows that a physician may respond to increased competition by greater effort at inducement. The marketplace model reveals that much, but not all, of physician inducement has corresponding behavior in ordinary business firms.

The evidence indicates that substantial variation must be affected by information flows but also by standard supply and demand variables. Removing all SAV variation that it is possible to remove may create improvements to social welfare, although it is not clear how much if any of social welfare this would explain.

Some independent issues close the chapter. Malpractice litigation has effects on both SID and SAV. Though payment by outcome is rare in physician practice, we observe it among traditional healers in Africa, where it is beneficial for patient cases that require the complementary effort of both physician and patient.

Summary

1 The SID models describe the physician as seeking to maximize utility over income, leisure, and inducement, and capture many observed physician behaviors.
2 SID includes the possibility of physicians deviating from their agency responsibilities to provide care to their self-interest rather than their patients' interests.
3 The target income hypothesis suggests that physicians use their discretionary advantage to achieve a target level of income or an improvement in their relative income.
4 In the McGuire/Pauly model, the physician faces trade-offs with income and leisure but also with income and the disutility of inducement.
5 Inducement behavior can be understood as the physicians having strong income effects in response to reduced practice profitability.
6 A marketplace model shows similar results with influences by ordinary businesses as it does with influences by physicians. Such influences may be harmful but need not be.
7 Small area variations (SAV) refer to the frequently wide inter-area and intra-area variations in the per capita use rates found for many medical and surgical procedures. Researchers have found such variations in the United States and many other countries.
8 The variations may be caused by imperfect information flows but also supply and demand factors.
9 SAV factors may harm social welfare but they need not.
10 Most of the evidence for the SAV is indirect, coming from studies showing the following:
   - Changes in practice patterns following physician education and monitoring.
   - Wide variations in utilization rates across small, homogeneous areas.
   - High, unexplained residuals in multiple regression analyses of inter-area utilization rates.
11 With SAV it is difficult to infer whether a particular area provides too much care or just the right amount.
12 Medical malpractice may represent issues of both SID and SAV. The impacts of potential malpractice liability on physician behavior vary by specialty.

Discussion Questions

1 Suppose that insurers monitored all health care payments to determine whether the services were appropriate. Would you expect to see more or less tendency toward SID?
2 Figure 15.2 shows how increased competition can lead to a higher degree of inducement at point E_2. This suggests that providers try to induce more usage to compensate for lower profit margins. Suppose, however, that the physician ends up at a different point, call it E'_2, where there is less inducement than at I_{E_1}. Could convex indifference curves be drawn so that the change to the m' rate of profit would lead to this reduced inducement?
3 If physician fees are fixed so that they do not adjust to changes in supply, explain how a firm whose demand curve slopes downward would react if the fixed fee were lowered, perhaps by the government.

4 What are some criticisms of the target income approach to SID? In the target income model, what determines the physician's target income? Would target incomes differ between physicians in practice? Would some physicians choose instead to be income maximizers?
5 In the utility-maximizing model, what forces limit a provider's ability and willingness to engage in SID? In the profit-maximizing model of SID, what are the costs to the physician of inducement? Give examples of inducement costs.
6 Assuming that SID is prevalent and substantial, what are the implications for policy? A policy to reduce Medicare payment rates? A policy to increase the number of medical school graduates?
7 What is SAV? What are some economic forces that can help explain SAV? What are some demographic and other considerations? How might physician uncertainty lead to SAV?
8 Do high utilization rates necessarily indicate the provision of unnecessary care? If not, why not?
9 If the cause of SAV is lack of information about the productivities of various procedures, would you expect SAV to have increased, decreased, or stayed the same over time? Why?
10 As new technologies become available more rapidly for given procedures, would you expect an increased or decreased amount of SAV?
11 Suppose large variations occur in use rates within a typical small area, as well as among small areas. What would this mean for policy?

Exercises

1 Explain why the indifference curves in Figure 15.2 are positively sloped.
2 In Figure 15.2, suppose that the initial profit is \$1 per unit of inducement ($m = 1$). Suppose then, that increased competition lowers m from 1 to 0.5.
   (a) Draw the new profit line $\pi = mQ_o + mI$.
   (b) Demonstrate the case where the change in profit increases the level of inducement.
   (c) Demonstrate the case where the change in profit decreases the level of inducement.
3 In comparing SAV among diseases and diagnoses, would more complicated diseases suggest greater or lesser variation?
4 Consider the approximation of the welfare loss due to inter-area deviations from the correct rate of care. All else equal, which procedures would yield the largest welfare losses—those with low price elasticities in absolute value or those with high price elasticities in absolute value? Why is this so?

Note

1 www.statelawyers.com/Practice/Practice_Detail.cfm/PracticeTypeID:63, accessed September 1, 2016.

Chapter 16

Health Care Labor Markets and Professional Training

In this chapter

- The Demand for and Supply of Health Care Labor
- Factor Productivity and Substitution among Factors
- Health Care Labor Supply and the Meaning of Shortages
- Medical Education Issues and the Question of Control
- Licensure and Monopoly Rents
- Other Physician Labor Issues
- Conclusions

Commensurate with its high share of GDP and the labor-intensive nature of its output, the health care economy employs a large number of workers. We use the term *labor* here in the general economic sense of production input that is distinct from capital and provided by human beings. Like other goods and services, the production of health services requires both labor and capital.

While many health sector workers are relatively unskilled, the health sector also requires large numbers of highly trained professionals. Most physician specialists obtain years of post-graduate education past medical school. This chapter addresses labor issues that range from general supply and demand principles applied to all health care occupations, to specialized topics involving shortages of doctors and registered nurses, medical education and licensure, and various practice decisions of physicians.

The Demand for and Supply of Health Care Labor

Chapter 1 described the magnitude and variety of health care occupations, and the importance of labor not only to the health sector, but to the overall economy. In 2014, 18 million people, representing 13 percent of total nonfarm employment, worked in the health services industries. These data do not include many other workers in the pharmaceutical and health insurance industries, as well as those in industries providing supplies, capital goods, and services for people providing direct patient care. Workers in some health-related occupations, such as pharmacists employed in drugstores, also are not included.

These numbers have increased substantially with the growth of the health economy. Between 1970 and 2012, the number of physicians tripled from 334,000 to just over 1 million; the number of registered nurses more than tripled from 750,000 to approximately 2.7 million.

How labor is used—and how it is combined with other factors of production—helps determine both the amount of health care provided and the wages and salaries of the providers. The productivity and training of health care providers are important to the working of labor markets and to the demands and supplies of labor.

We begin by describing the determinants of labor demand. We derive demand for a factor of production, either labor or capital, from the demand for health. We demand health care providers because we demand health care, and we, in turn, demand health care because we demand health.

Production Functions and Isoquants

Recall that the production function describes the relationship of factors of production (the inputs) to the resulting goods and services produced (the outputs). Under the existing technology and know-how, it shows the maximum sustainable output obtained from all possible combinations of inputs, such as labor, materials, buildings, and equipment.

Economists often simplify the production relationship as follows:

$$Q = f(L, K) \tag{16.1}$$

where Q represents output over a period of time, and L and K represent the quantities of labor and capital inputs over the period. We can illustrate many features of a production function graphically through isoquants. Recall also that an isoquant represents all combinations of inputs (e.g., labor and capital) that result in a given level of output. Figure 16.1 (panels A and B) describes isoquants for two different situations.

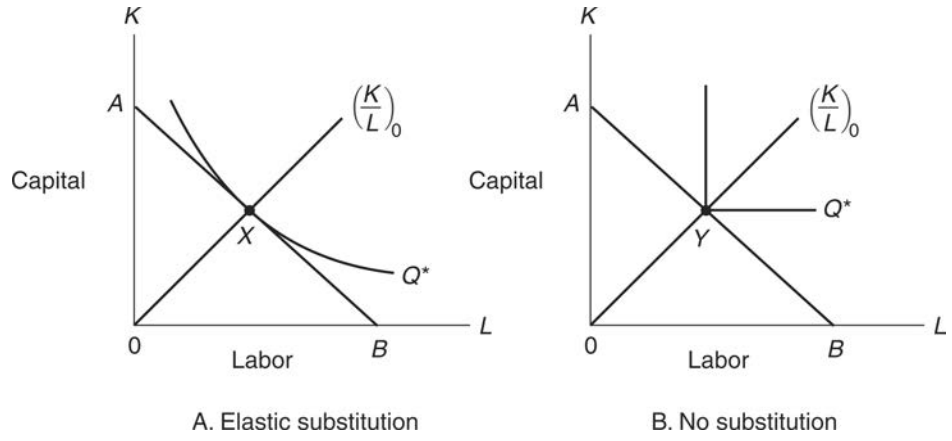


Figure 16.1 Production with Varying Rates of Substitution

In panel A, the isoquant shows a technology in which labor and capital are fairly good substitutes for each other so that labor and capital can be combined in many different proportions to produce output. The budget line, *AB*, reflects the trade-off between capital and labor, and point *X* is the location at which the costs of producing Q^* units of output are minimized.

With the given set of input prices (wages and payments to capital) the optimum (cost-minimizing) capital–labor ratio $(K/L)_0$ is the slope of a ray from the origin through point *X*. Due to the curvature of the isoquant, a change in relative input prices changes the optimum capital–labor ratio. In panel A, labor and capital are good substitutes as we move along a given isoquant. Lower prices of labor (i.e., a flatter budget line) will lead to a relatively large substitution of labor for capital, and vice versa.

In contrast, panel B shows a technology in which labor and capital are not good substitutes; as drawn, they must be used in fixed proportions to one another. Although point *Y* represents the same ratio (*K/L*) as point *X*, changes in the factor prices will not change the capital–labor ratio. Specialized surgeries, for example, may require specific ratios of labor to capital with little substitution available. The degree to which substitution among inputs is possible, either between health care labor and capital, or among different types of health care labor, is a key issue in health resource planning and in determining the efficiency of production exhibited by health care firms.

The demand for any type of health care labor depends in part on these substitution possibilities. The demand for a factor of production also depends on the price of the output. Consider an example. Suppose you were working as a skilled worker in a bicycle manufacturing plant, and suppose that bicycle riding was becoming more popular. The increase in demand would result, at least temporarily, in a higher price for bicycles. More bicycle workers would be demanded as a consequence.

Marginal Productivity of Labor

Consider that the demand for an input, and consequently the wage paid to the input, will depend at least in part on the input's productivity. This is one explanation why college-educated workers earn more money than others. If college-educated workers are more

productive than others, then the demand for them will be greater. Although this is only one theory that explains the earnings advantage of college graduates, it illustrates the premise that more productive laborers are in greater demand.

Reconsider the idea of marginal product—the increase in output when an input is increased by one unit. Suppose that labor in equation (16.1) represents the number of laboratory technicians employed. If we add one lab technician to the production process, holding the available lab equipment and materials (i.e., K) constant, we expect to increase lab output. This increase is the marginal product of lab technicians.

Although additional lab technicians in the production process will likely increase total output, they likely will add incrementally less output. As we increase an input, holding all others constant, output will tend to increase but at a decreasing rate. This illustrates the law of diminishing returns.

The number of lab technicians to hire depends also on the price of output. Suppose that lab tests sell for $100 per test. Suppose also that an additional technician would increase output by four tests per day. If these tests sell for $100 each, the technician is bringing in an additional $400 in revenue. The extra revenue generated is called the marginal revenue product (*MRP*).

Would it pay to hire this extra technician? Clearly the answer depends on the wage per day. If technicians earn $150 per day, the technician nets the hospital a $250 gain ($400 marginal revenue product less $150 wage), so it pays to hire another one. It always pays to hire laborers whose marginal revenue products exceed their wage.

Would it pay to hire still another technician with a marginal product of three tests per day? Because output (the tests) sells for $100 each, this next technician is netting the hospital $150 because the marginal revenue product, $300 in this case, exceeds the wage, $150. It will pay the firm to continue to hire more workers up until the point where the marginal revenue product equals the wage.

Figure 16.2 represents the marginal revenue product curve for lab technicians. The curve slopes downward, reflecting the law of diminishing returns. In competitive markets, we find the *MRP* curve in Figure 16.2 by multiplying the marginal product curve (not shown) by the price of output. The optimal number of lab technicians depends on the wage rate. At wage W_1, the optimal input demand at point A is L_1. At a higher wage, W_2, the firm would demand fewer technicians, L_2, at point B.

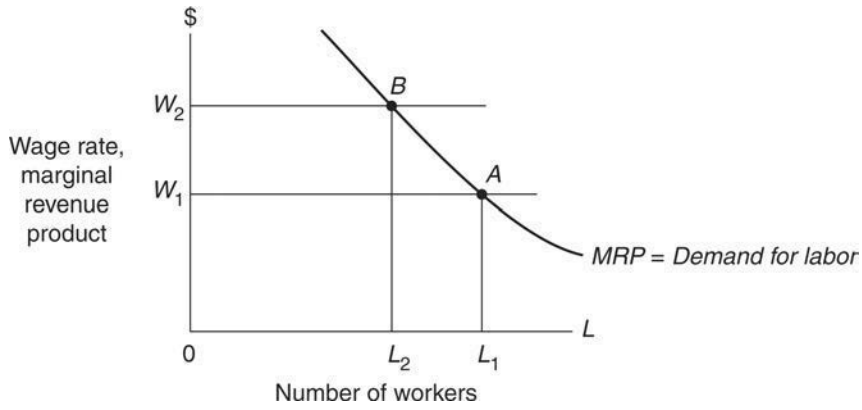


Figure 16.2 Firm Hiring Decisions at Different Wage Levels

This analysis shows that the demand for labor is precisely the marginal revenue product of the labor curve. Through the marginal product, this curve is closely related to the production function for the laboratory test. It is also directly related to the price of laboratory tests. If their price increases, the marginal revenue product increases, and more labor is demanded. If better machines make the lab technicians more productive, the demand for their services will also increase.

Factor Substitution and Labor Demand

At this point, recall the meaning and importance of the substitutability of one input for another. Suppose, for example, a newly invented machine allows lab technicians to perform certain functions previously performed by radiologists. This makes technicians better substitutes for radiologists. As a result, the demand for technicians will tend to increase. This change will also tend to shift the demand for radiologists and probably make it more elastic—that is, flatter and more responsive to their wage rate. As a result, if the firm finds it can substitute more easily between inputs, it will become more resistant to input price changes, replacing increasingly expensive inputs with cheaper substitutes.

In recent decades, firms and policymakers have looked closely at substitution possibilities in their efforts to control health care costs. To the extent that nonphysician providers can substitute for physicians, firms can hire lower-cost workers, hence cutting costs and possibly increasing the amount of the service provided. We address this issue in more detail later in the chapter.

We find the market demand for various occupations by horizontally adding the demands of the individual firms. The market demand for laboratory technicians in Figure 16.3 is the downward-sloping curve labeled D_1. Under competitive conditions, the labor market equilibrium, and consequently the equilibrium wage, depend on the interaction of both demand and supply.

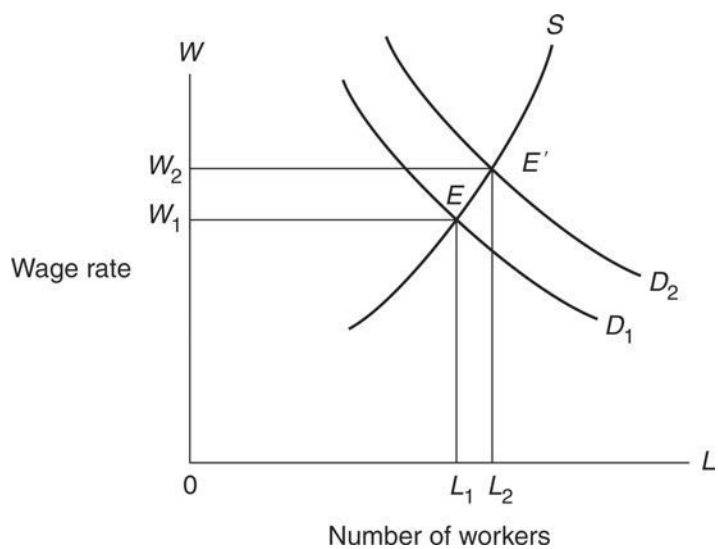


Figure 16.3 Market Demand and Supply for Laboratory Technicians

The Supply of Labor

The supply of labor tends to slope upward, implying that a higher wage rate in a given market will attract more workers or labor time. First, those workers currently employed may choose to work more hours if higher wages are offered. Second, similar workers may be attracted into the market from elsewhere. We show the labor supply curve for lab technicians in Figure 16.3 as *S*.

The equilibrium market wage, W_1, for technicians, along with the market-clearing number of technicians, L_1, occurs at the intersection of the supply and demand curves. Consider that the market equilibrium wage will tend to increase in response to any demand-increasing event. These events include increases in the firms' desires to substitute lab technicians for other laborers, increased productivity of the technicians, and increases in the price of the lab outputs. Here, demand increases to D_2, increasing the equilibrium wage to W_2 and employment to L_2, at point E'.

Similarly, the market wage will tend to fall in response to events that increase the labor supply. These include increased graduations of trained technicians or influxes of technicians into the market from other professions or other locations.

Changes in supply can take place relatively quickly for those health care occupations requiring minimal education or training. For physician specialists and others requiring the highest levels of education and training, the number of new professionals is determined by the admission decisions of medical schools as well as the decisions of applicants, both made many years earlier. Physicians must be licensed by a state in order to practice in that state. Requirements for licensure include graduation from an accredited medical school, passing a licensure examination, and completing one to two years of internship or residency in an accredited graduate medical education program. Many graduates, nevertheless, complete three- to four-year residency programs. Many physicians also become board-certified specialists. The requirements typically include advanced residency training for three to six years, practice in the specialty, and passing the board examination.

Through education and training, medical students make investments in their human capital. Economists treat the decision to invest in human capital with the same tools used to analyze investments made by businesses in physical capital. The decision maker will consider the revenues associated with investment along with all costs, including any opportunity costs. For medical students, the forgone earnings associated with the time it takes to complete their medical education is an important opportunity cost, but monetary values must also be imputed to nonpecuniary gains, such as the satisfaction from helping the ill and the prestige associated with being a physician. The internal rate of return is the rate that equates the present value of revenues with the present value of costs (see the Appendix to Chapter 4 for a discussion of discounting). Prospective medical students, in principle, will compare the return from medical education with those of other possible occupational choices and select the highest one.

Information regarding the rate of return to a medical education also is useful to the policy analyst. Suppose we discovered that the average rate of return to medical education is high and rising. This information might mean that physicians are becoming increasingly scarce, suggesting a shortage. Numerous studies have attempted to estimate the rate of return to a medical education, and we will describe some of them later in this chapter.

Factor Productivity and Substitution among Factors

The supply of health services, and consequently health care prices, depends on the number of workers. However, productivity of labor inputs represents a critical determinant of

supply. Productivity improvements increase output or, alternatively, the same output can be produced with fewer inputs. In a macroeconomic sense, general improvements in productivity represent a major source of economic growth and rising standards of living. At a microeconomic level, productivity gains in an industry can lead to lower prices for the goods and services produced in that industry and possibly to higher rates of remuneration for workers.

The productivity of a factor of production can be measured as the average product of the factor—that is, the ratio of total output (Q) to the amount of a particular labor input (L):

$$\text{Average product of labor} = Q/L \tag{16.2}$$

This definition corresponds to the concept of average product used in microeconomics and is distinguished from marginal product, which we define as the change in output associated with a one-unit increase in the input, holding all others constant.

Despite the simplicity of the concept of average product, difficult problems occur in measuring it. The product may be heterogeneous, consisting of many different outputs. In such cases, one often uses the dollar value of output for the numerator, Q. Similarly, many kinds of labor are used in the production process. In such cases, a weighted sum of related inputs is often used in the denominator term, L.

Measurement of Physician Productivity

Reinhardt (1972) undertook a classic study of physician productivity. He examined general practitioners in private practice for three measures of output: total patient visits, office visits, and patient billings. In addition to physician time, he considered the use of various auxiliary personnel. Reinhardt estimated the marginal product of physician time—the increment to output resulting from the addition of one hour of physician time to the production process.

He found that the marginal product tended to increase up to where the physician was working a total of about 25 hours per week; the marginal product eventually declined to zero at about 110 hours per week. He found that starting from a base of 60 hours per week, a 1.0 percent increase of physician input would result in an increase of 0.8 percent in the number of patient visits produced.

In addition to studying the physician's productivity, Reinhardt examined the substitution possibilities between physician and other labor inputs. Consider physician aides. The aides' marginal products were highest when about one aide was present per physician. Physicians could improve productivity of their practices and increase profits if they doubled the number of aides from two aides per physician (the average he found) to four aides per physician.

The Efficient Utilization of Physician Assistants: Substitution among Inputs

The possibility that physicians were underutilizing aides was a provocative one. In 1988, Brown refined Reinhardt's work and found, for example, that an additional dollar spent on hiring more practical nurses generated more output (office visits) than an additional dollar spent on physician inputs. The conclusion! Physician practices would become more profitable if one substituted practical nurses for physicians.

In addition, Brown estimated that physicians in group practices were on average 22 percent more productive than those in solo practices. He suggested that this figure, much higher than the 5 percent estimated by Reinhardt, resulted from advantages that group practices have in employing physician assistants.

Other research provides considerable evidence on the substitutability between physician time and other labor inputs. Escarce and Pauly (1998) found that each hour of time for an office-based internist substitutes for $60 in nonphysician costs or vice versa. Elsewhere, the growth of managed care organizations has heightened interest in PAs and nurse practitioners (NPs) to improve productivity and lower costs. Jacobson and colleagues (1998/1999) report that PAs/NPs can perform 50 to 90 percent of the tasks of primary care physicians without compromising quality when they work collaboratively with physicians. They also found that PAs/NPs have greater scope of practice and autonomy as the proportion of managed care patients in a health care organization increases. Despite legal impediments and other limits on the use of PAs/NPs, the delivery of primary care is likely to rely increasingly on these skilled physician substitutes as organizations try to move toward their optimal input mix. Box 16.1 summarizes some of the most recent work on productivity.

Health Care Labor Supply and the Meaning of Shortages

We have seen that the total number of health care professionals has increased, substantially. The population of ordinary citizens (the potential consumers) has also increased. Table 1.3 made clear that the number of professionals has been increasing more rapidly so the number of professionals per 100,000 population also has increased. As noted previously, the rate of increase for physicians and registered nurses per capita has been sharp. Table 16.1 provides more detail on physician practices.

BOX 16.1

Recent Productivity Studies

The basics of productivity theory outlined above follows the standard textbook model in which a production function shows the maximum sustainable output associated with any given level of inputs. In theory, the inputs are homogeneous and the production function should not be affected by financial or other incentives. It is purely a technological phenomenon. Two recent contributions highlight some of the nuances of dealing with productivity for health care services where both outputs and inputs are difficult to measure.

Kantarevic and colleagues (2011) compared various measures of productivity for primary care physicians in the Canadian province of Ontario under two different payment systems. One was traditional fee-for-service (FFS); the other was an enhanced FFS model, introduced in 2003, that provides financial rewards for improving quality and access through preventive care; managing chronic diseases; and providing after-hours services. The study found that those physicians who joined the enhanced FFS plan had significantly higher productivity (by about 6–10 percent) than comparable

FFS physicians in terms of services, visits, and number of patients. However, a substantial portion of these gains was attributed to an increase in the supply of physician labor rather than increases in output for any given work effort.

Elsewhere, Doyle, Ewer, and Wagner (2010) take advantage of a natural experiment in a large U.S. Veterans Administration (VA) hospital which randomly assigned 30,000 patients to clinical teams from two academic affiliates. The randomization ensures that patients have very similar conditions and characteristics. One of the affiliates is considered a top medical school while the other has a lower rating. The teams had access to the same VA facilities, nursing staff, and other support staff.

Although patient health outcomes were essentially the same, there were large differences in treatment costs. Teams from the highly ranked affiliate incurred considerably lower costs—10 percent overall and up to 25 percent for more complex conditions. The differences were attributable to higher use of diagnostic testing, with correspondingly higher costs, by teams from the *lower* rated affiliate. The study indicates that high-quality physician teams can be much more productive, i.e., attain the same output with fewer total inputs, than lower-quality teams. Physician quality matters a great deal and this realization has significant policy implications relating to standards at medical schools and hospitals.

Despite these increases, a recurrent policy concern has been the availability of various critical categories of professionals. The issue usually hangs on whether we have or will have enough of them; that is, whether there will be shortages or surpluses.

Availability of Physicians

Occupational analysis often focuses on the availability of personnel to provide needed or demanded patient care. Many physicians, including researchers and administrators, do not provide care for patients. Because the portion providing office-based patient care to the public has not changed much over recent decades, the rapid increase in physicians per capita also has meant a greater number of patient care physicians per capita.

Of the 826,000 active medical doctors in the United States in 2012, 785,000 provided direct patient care (Table 16.1). Of those, three-fourths were office-based and the rest were hospital-based, including residents. As we noted previously, the pattern of the portion of physicians available for patient care has not been changing rapidly.[1]

Physicians form a large number of specialties rather than a homogeneous group. About 50 percent of office-based physicians practice in primary care (general and family practice, internal medicine, pediatrics, and obstetrics/gynecology), and 20 percent are in general surgery or the surgical specialties. The remainder have other specialties such as dermatology (skin), rheumatology (joints), or oncology (cancer). Analysts have long expressed concerns about specialization as well as about uneven distributions between rural and urban areas in the United States.

Planners and policymakers often worry about having adequate quantities of workers and avoiding serious shortages, especially of physicians. In the mid-twentieth century, a need-based method, based entirely on medical considerations, became the dominant approach to determining physician requirements. This approach, illustrated by the classic study of Lee and Jones (1933), calculated the number of physicians required to serve in a given market area from the needed number of procedures that in turn related to the incidence of morbidity (illness) in the population. Under what Fuchs (1974) has criticized as the monotechnic approach, health services planners assumed that a single technique that includes a fixed

Table 16.1 Physicians by Type of Practice: 1975–2012

	Number in Thousands				
	1975	*1985*	*1995*	*2005*	*2012*
All Physicians	394	553	720	902	1,027
Professionally Active	340	497	625	762	826
Nonfederal	312	476	604	NA	NA
Patient Care	288	432	564	718	785
Office-Based	213	329	427	563	586
Hospital-Based	75	102	137	155	199
Other Active	24	44	40	44	41
Federal	28	21	20	NA	NA
Inactive/Unclassified/ Unknown Address	53	56	95	140	201

Note: NA = Not available.
Source: U.S. Department of Commerce, *Health United States, 2015* (Table 93) and earlier issues.

amount of physician time is required to treat each particular type of illness. Total physician time, or physician equivalents, was determined by aggregating over a broad range of medical conditions. Economic considerations, such as the potential substitution of other inputs for physician time, changes in technology, differences in patients' preferences, and even costs were generally ignored. Box 16.2 illustrates such limited thinking in connection with perceived shortages of primary care physicians.

Economic Definitions of Shortages of Health Professionals

Economic definitions of labor shortages usually differ from those based solely on medical grounds. Economists apply definitions based on considerations of how characteristics of a given market for professionals deviate from those found in an ideal, highly competitive market. As a result, several approaches for determining shortages occur in the literature.

BOX 16.2

Dealing with Shortages of Primary Care Physicians

The aging of the population, an increased emphasis on prevention and wellness, and growth in those with insurance as a result of the ACA are expected to drive major increases in the demand for primary care. At the same time, modest increases in medical residency slots, high physician retirement rates, and the continued preference for specialty practice among medical school graduates have fueled predictions of shortages of primary care physicians. Even with increases in medical school enrollments, as

described later in this chapter, and various ACA provisions to increase primary care supply, some analysts predict substantial physician shortages by 2025.

Fortunately, many recognize the implicit assumption in these forecasts of a fixed-proportion production function, i.e., of the monotechnic approach described above. Auerbach and colleagues (2013, p. 1993) highlight the fundamental limitation of such forecasts:

> [T]hey implicitly assume that the number of full-time equivalent primary-care physicians available today is, on average the optimal amount needed for a given population and that, with slight adjustments for factors such as population aging, this amount will not change appreciably in the future.

The growth of convenient ambulatory care centers, typically walk-in clinics and urgent care centers, that are much less labor and capital intensive than traditional facilities can increase the availability of relatively low cost care. However, much of the interest in closing the gap between demand and supply calls for an expanded role for other health professionals. These include nurse practitioners and physician assistants as well as other clinicians such as chiropractors, acupuncturists, and pharmacists who can help deliver substantial amounts of primary care. Even the patient's role in wellness management cannot be overlooked. Above all, there are new models of team-based care that take advantage of the wide range of health care workers and new technology (e.g., telemedicine and quality dashboards). If the promising results of these models can be reproduced on a wider scale, the projected crisis may turn out to be less serious than predicted.

Sources: Aurbach et al. (2013); Dill et al. (2013); Bodenheimer and Smith (2013); Green, Savin, and Lu (2013); and Chang, Brundage, and Chokshi (2015).

EXCESS DEMAND Begin with the conventional economic definition of a shortage: the excess of the quantity demanded over the quantity supplied at market prices. Figure 16.4 depicts a shortage defined in this way: The labor shortage at the wage W_1 is equal to $L_{d1} - L_{s1}$. If the wage instead had been W_2, no shortage would exist. This is because at the higher wage, less labor services would have been demanded and more would have been supplied.

This definition, though valid, raises critical questions in the case of health care workers. Why didn't the wage rise to equilibrium, thus automatically eliminating the shortage? The usual case of persistent excess demand is associated with stickiness in wages or prices imposed by law or regulation. A common example in some American and European cities is legally enforced rent control in the housing market.

What would cause wage stickiness in health care labor markets? It seems doubtful that health workers' wages are sticky in the sense of administered rents or prices, or that shortages in terms of excess demand are a serious policy problem. Some analysts have argued that shortages due to unmet demand are not serious concerns for most categories of professionals.

RELATIVELY RAPID INCREASES IN WAGES: DYNAMIC SHORTAGES An unnecessary focus on excess demand also obscures the fact that economically meaningful shortages of professionals may well exist even when supply and demand are in short-term equilibrium. In particular, a shortage may occur when demand and supply conditions change over time. Suppose, for example, that demand for a category of health professionals expands over time,

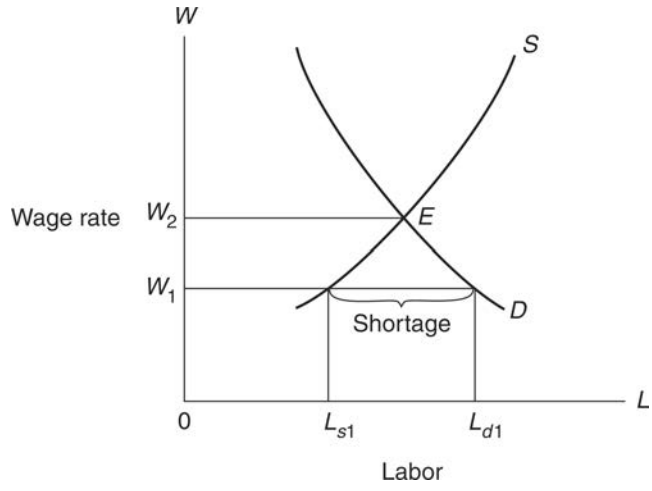


Figure 16.4 An Economic Shortage

and that the supply of these professionals is slow to respond or even perhaps faces barriers in responding. The result may be a large rise in wages relative to the wage gains of other professionals.

To illustrate, consider Figure 16.5, which depicts the demand for physicians (or specialists, such as highly skilled surgeons) at two points in time: an initial Period 1 and a subsequent Period 2. We compare the equilibrium quantity in Period 1 with that in Period 2. The wage increase from W_1 to W_2 may indicate a shortage, even though quantity supplied equals quantity demanded (at L_1 and L_2) in both periods. We would say a shortage exists if the relative wage of highly skilled surgeons has risen sharply relative to that of other professionals. The measure of shortage under this approach is the relative wage and the direction of its movement.

Several variations on this general approach have been described. The pattern of wages over time may be more complex than the movement described from equilibrium at W_1 to equilibrium at W_2. We might find, for example, that the initial market response to increased demand for the professionals would be to raise wages to W_2. Only after these high wages had induced the expansion of supply to S_2 would we observe market wages adjusting to W_3 (and quantity supplied to L_3). Under this scenario, the professional's wage for a time falls, here from W_2 to W_3. Thus, a decline in relative wage during a given period may reflect a long-run adjustment offsetting a shortage and not necessarily an indication of excess supply.

RELATIVE RATES OF RETURN How should we measure the monetary gains from professional training? Hansen (1964) provides a classic measurement approach that is both plausible and consistent with theory. He proposes that the relevant measure of monetary gains to a given health professional group must take into account the various opportunity costs incurred by professionals in obtaining their training using the internal rate of return.

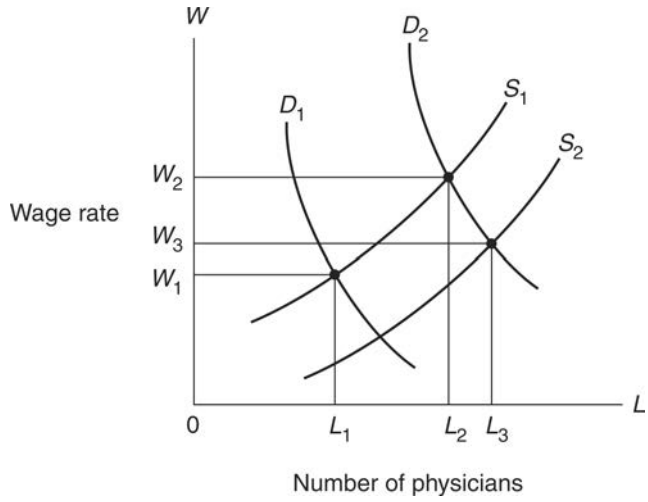


Figure 16.5 Changes in Equilibrium over Time Depending on Supply Adjustments

The internal rate of return is the discount rate that equates the present value of the stream of costs to the stream of revenues from education. The higher the rate of return, the greater the financial rewards are to investment in the human capital attained through education. To determine whether a given health professional group is in relatively short supply, we can compare the rate of return to that of other professionals and examine these comparative data over time.

High, even excessive, rates of return may occur whenever the supply of labor fails to respond quickly to changes in demand. In some instances, the underlying reason for this slowness in response may be barriers to entry faced by potential health professionals. A barrier to entry exists, in this case, when a potential health professional faces higher entry costs than incumbents faced.

In some cases, potential entrants are completely barred from entry. Such barriers would occur if controls on slots in health professional schools limited entry. They also occur to some degree whenever entry to the profession is limited by licensure laws. The issue of licensure laws is of special interest here because it is common in the health professions. Because of this connection, we treat the empirical literature on rates of return to physician education in the discussion of licensure in a later section of this chapter.

The Role of Monopsony Power: Shortages of Registered Nurses

Practitioners who describe health care labor availability often rely on reported percentages of unfilled, budgeted positions. One explanation of this measure (using Figure 16.4) is excess demand. Excess demand, however, is generally temporary; as long as prices are not rigid, price increases will tend to cure the problem. More plausibly, the analyst will focus on

changes in the percentage of unfilled, budgeted positions, analogous to dynamic definitions of shortages. First, however, we must address the problem of interpreting reported data for monopsonistic markets. A monopsony is a market that in theory has only one buyer; for example, a monopsony would be one hospital that hires virtually all registered nurses in the market. In Canada's single-payer health system, individual provinces have monopsonistic market power in paying hospitals and professionals.

MONOPSONISTIC LABOR MARKETS Under monopsony, a clinic or hospital may report unfilled, budgeted positions, for example, for registered nurses, even when the firm is actually in equilibrium. The paradox is that such a monopsony firm may announce that it wishes to hire more nurses even though it is unlikely to take the necessary steps to do so.

The paradox is explained by the monopsony firm's upward-sloping supply curve. Because it is a big employer, it has the power to influence nurses' wages and, thus, to induce more nurses to work by raising the average nurse wage level. The monopsony clinic or hospital is willing to hire more nurses at the current wage, but it has no intention of paying a higher wage in order to hire more nurses.

These ideas can be illustrated with the help of Figure 16.6. Imagine that one hospital is the only demander of nurse labor in the market. The hospital's demand curve for nurse labor, labeled *D*, represents the marginal revenue product curve for nurses employed at that hospital.

Under monopsony the supply curve for labor will no longer represent the marginal labor cost, *MLC*, to this hospital. If this hospital was a competitive hirer, competition would have meant that the hospital could have hired as many nurses as it wanted at any given wage. In a competitive market, N_d workers will be hired at wage W_1.

Consider the monopsonist's marginal labor cost curve, *MLC*. A monopsonist hospital seeking to add one nurse to its labor force must pay a higher wage than before in order to induce this marginal (extra) nurse to work. But it then must pay all of its employed nurses a higher wage. The problem arises entirely because the labor supply curve is rising, as is typical under monopsony, instead of being flat as is typical under competition.

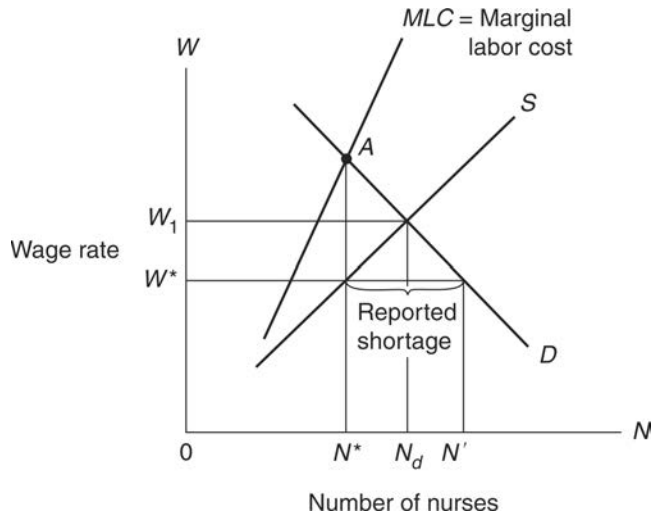


Figure 16.6 Nurse Shortage under Monopsony Conditions

The extra (marginal) labor cost incurred for that one nurse is not just the wage it pays, but also includes the extra wages it must pay all its other nurses. As a result, the marginal labor cost curve, *MLC*, will lie above the labor supply curve, *S*. The monopsonist will hire fewer nurses and pay a lower wage than will a competitive firm.

Early evidence found only tenuous support for the classical monopsony framework so scholarly interest turned to a nuanced "new monopsony" model that featured worker-attachment to hospitals from sources other than hospital concentration. That is, in this new approach, positively-sloped labor supply curves are not necessarily the result of monopsony power. Hirsch and Schumacher (2005) found very modest decreases in RN wages in response to increases in hospital concentration. They argue that this provides some support for classical monopsony in the short run. But they also show this effect cannot be sustained over the long run due to the relatively high mobility of nurses among employers. They conclude that "absent more compelling evidence, nursing should not be held up as a prototypical example of monopsony—classic or new" (p. 969).

REPORTED SHORTAGES Suppose now that the hospital acts to maximize its profits. It would find it profitable to hire an additional nurse as long as the extra revenue the new nurse brought in, the marginal revenue product, exceeds the extra cost of the nurse to the hospital, the marginal labor cost. Thus, as long as the demand curve (the marginal revenue product) lies above the marginal labor cost curve for a given level of nurse employment, it always will pay to hire more nurses. The hospital achieves its profit-maximizing complement of nurses where $D = MLC$, an employment level of N^* nurses in Figure 16.6. The equilibrium wage is found on the corresponding point of the labor supply curve; here it is W^*.

At this equilibrium wage, W^*, the hospital would desire to hire N' nurses. It may well budget for these nurses and effectively report a shortage of $(N' - N^*)$. The hospital acts as if it did not realize that if it wants more nurses to work, it must pay its nurses more. At any rate, a shortage in this case has a limited meaning; it only means additional nurses are desired at the current wage level.

This theoretical point suggests the hazards of interpreting data on unfilled budget positions. The possibility of monopsony power in labor markets suggests that unfilled positions data may overstate the problems of nursing availability. The problem is somewhat mitigated by the fact that these data typically represent unfilled budgeted positions because a hospital is unlikely to budget for all the positions it might desire at the current wage. In any case, the analyst might wish to examine the percentage of unfilled budgeted positions. If the number of unfilled positions rises rapidly, it may suggest an increasingly relative scarcity of nurses.

Medical Education Issues and the Question of Control

Most other health care workers carry out their tasks under the direction of physicians, and from their authority in treatment decisions, physicians are the dominant providers in the health economy. As a result, economists have concentrated on the training and practice of physicians even though they represent a minority of patient care providers. Many professions require a considerable length of time for education and training, but the time period for formal training of physicians is among the longest. In addition, medical education poses the question of who has control. Does the medical profession itself exercise control over access to medical

education in order to improve its own profitability? It is helpful to study medical schools and their funding, as well as information about the supply of potential medical students.

Sources of Medical School Revenues

In the United States, as elsewhere, the government heavily subsidizes medical school education. This situation comes from public concern for the adequacy of the supply of physician labor. The rapid growth in medical school enrollments did not arise by accident. Several actions by Congress, beginning with the Health Professions Educational Assistance Act in 1963, provided grants to medical schools and financial assistance to students. Some of this federal support hinged on enrollment increases. In 1971, however, federal support to medical schools increased substantially and came in the form of capitation grants, which rewarded the medical schools for expanding their enrollments by giving money on a per-student basis, initially about $3,000 per year per student. With fears of physician surpluses emerging by the late 1970s, assistance under the Health Professions Educational Assistance Act was phased out, and enrollments stabilized.

Nevertheless, medical education is still subsidized heavily. Tuition represents a relatively small source of revenues for many medical schools so the student pays only a small portion of the true cost of the investment in education. Governmental support for operating revenues and through grants and contracts is typically about 30 percent of total revenues. The largest share (about 50 percent) comes from reimbursements for health services provided to patients (Jones et al., 1998).

Teaching Hospitals, Medical Schools, and Joint Production

Medical education is a good example of joint production. That is, medical schools produce at least three products jointly:

- Medical education.
- Patient care.
- Research.

To reimburse for patient care or to fund medical education appropriately, it is necessary to determine the pure costs and the joint costs of these activities. An example taken from Newhouse (1978) illustrates these terms.

In Table 16.2, the total annual cost for a medical school that produces only education and patient care is shown to be $60 million. If the school produced only education with only the minimum patient care needed to do this, its costs would be $50 million. If it produced only its present volume of patient care and no medical education, its costs would be $30 million.

Incrementally, the cost of patient care raises the school's budget from $50 million to $60 million. Thus, the *pure* cost of patient care is the extra $10 million. Reasoning in a similar fashion, adding education to the cost of patient care raises the budget from $30 million to $60 million. Thus, the pure cost of education is $30 million.

Notice that the difference between total cost of this hypothetical medical school and all the pure costs is $20 million. This $20 million is called the joint cost. It follows that if the school were reimbursed only for pure costs, it would run a deficit. Much of the controversy with respect to funding revolves around the problem of who will pay for the joint costs.

The issue of joint production has centered on the teaching hospital, which also jointly produces patient care and graduate medical education (GME) by providing residency and

Table 16.2 Hypothetical Example of Joint Production at a Medical School

	(In millions of dollars)
Total cost of school	60
Cost if school produced only patient care	30
Cost if school produced only education	50
"Pure" cost of education[a]	30
"Pure" cost of patient care[a]	10
Joint costs[a]	20

Notes: [a] The pure cost of education is total cost (60) less the cost of producing only patient care (30). The pure cost of patient care is total cost (60) less the cost of producing only education (50). Joint costs are total costs (60) less all pure costs.

Source: Based on information from Newhouse (1978).

medical research. In particular, with the substantial cost differences between teaching and nonteaching hospitals, about 20 percent according to some estimates, third-party payers are concerned about whether they are implicitly subsidizing GME. Medicare, which provides most of the explicit funding for GME—$10 billion in 2012 or about $110,000 annually for each of the 90,000 residents that Medicare supports under a cap that was established in 1997—is also concerned about the lack of accountability in the existing payment system. In 2010, the Medicare Payment Advisory Commission (MedPAC, 2010) recommended a major overhaul of GME funding. It would establish a performance-based incentive structure where programs have financial incentives to prepare graduates with the skills needed to improve quality of care while helping to contain costs.

Foreign Medical School Graduates

Physician supply in the United States depends to a significant degree on foreign medical school graduates (FMGs), and reliance on them continues to grow. As a proportion of the total number of active physicians, FMGs (excluding Canadian) increased from less than 14 percent of the total in 1963 to 27 percent in 2013. Critics often argue that the United States and other rich nations drain valuable talent from many poor countries that have inadequate health care systems.

Nevertheless, the availability of physicians from other countries can have important policy and planning implications. Foreign national FMGs can increase the responsiveness of physician supply in the United States to changes in the physician wage.

Rapid increases in physician wages send a market signal to potential physicians, increasing the estimated rate of return to an investment in medical education. It takes a long time, however, for new applicants among American college graduates to get to medical school, get training, and enter practice. As a result, the supply of new American physicians will respond slowly to the wage signal. Foreign national FMGs, already trained but currently practicing elsewhere, can respond more quickly so the availability of foreign national FMGs makes total physician supply in the United States, in principle, more elastic.

Policymakers can apply these facts about FMGs during periods of shortages and increasing fees. Immigration policy can be relaxed to admit more foreign national FMGs during periods of physician shortages, and vice versa during periods of perceived physician surpluses. This happened during the 1960s, when preferential treatment was given to foreign national FMGs. Requirements subsequently were tightened by the 1976 Health Professions Educational Assistance Act once shortages were no longer perceived.

In addition to immigration policy, the number of FMGs, including U.S. graduates of foreign medical schools, reflects certification and licensure requirements. All FMGs must take certification examinations required for admission into an approved graduate medical education program. The 1976 act also placed restrictions on the access of FMGs to graduate medical education.

Advocates for tightening standards for FMGs usually claim that these graduates are inferior in quality to those educated in U.S. and Canadian medical schools. The claims are based on comparisons of examination performance and other measures of the credentials and personal attributes of FMGs and their U.S. counterparts.

Studies addressing the issue, however, find little difference between FMGs and domestically trained physicians. Some have argued that differences will more likely arise in ambulatory care settings because there is less organizational control than in hospital settings. A study that examined more than 14,000 patient episodes by nearly 1,200 physicians in three specialty groups found little difference in performance. FMGs, in fact, sometimes outperformed U.S. medical school graduates (Rhee et al., 1986).

The Control of Medical Education

In 1974, Victor Fuchs wrote that "most economists believe that part [of physicians' high incomes] represents a 'monopoly' return to physicians resulting from restrictions on entry to the profession and other barriers to competition" (p. 58). Fuchs refers to the claim that physicians restrict entry to their profession in order to drive up prices for their services and make larger incomes for themselves.

Do physicians control entry to their profession in order to earn above-normal returns on their investment? To answer this question, we first ask whether physicians do, in fact, earn above-normal returns. Historically, physicians often earned above-normal returns. Distinguished economists found this to be the case using data from early in this century, and believed that control of entry was the cause. Subsequent studies found high returns in at least some more recent historical periods.

We must further ask how physicians can control entry. Our present ideas and beliefs about the role of organized medicine in controlling entry owe much to Kessel (1958), who argued that organized medicine attained monopoly power through the licensure of physicians and the control of access to medical education. The first, licensure, is explored later in this chapter. At present, consider Kessel's account of the control over medical education exerted by physicians primarily through the American Medical Association (AMA).

Control over Entry

Shortly after the founding of the AMA in 1847, the organization campaigned state by state to get the medical profession controlled through licensure. Having largely achieved this goal by the turn of the century, the AMA turned its attention to the control of medical schools,

which had proliferated in number. In 1906, the Council on Medical Education of the AMA inspected the 160 medical schools existing at that time, declaring only slightly more than half of them to be acceptable, many with low admission standards, poor laboratory facilities (insufficiency or absence of microscopes), and minimal exposure to clinical material.[2] The council sought support for this position through the Carnegie Foundation, which in 1910 issued the Flexner Report calling for substantial reductions in the number of medical schools and control on their quality. Following this report, the number of medical schools fell to 85 by 1920 and to 69 by 1944.

In examining the impact of the Flexner Report, Kessel argued:

> If impact on public policy is the criterion of importance, the Flexner Report must be regarded as one of the most important reports ever written. It convinced legislators that only the graduates of first class medical schools ought to be permitted to practice medicine and led to the delegation to the AMA of the task of determining what was and what was not a first class medical school.
>
> (p. 28)

He likened giving the AMA charge over determining the supply of physicians to "giving the American Iron and Steel Institute the power to determine the output of steel" (p. 29). The AMA also was able to gain control over the internship/residency process through its ability to certify hospitals for such training. It also maintained control over the process through which physicians become board-certified. The picture is one of significant power and means to control entry.

The AMA also was able to exercise control over substitute providers (e.g., optometrists, podiatrists, chiropractors) by influencing licensure to limit their scope of practice and later to limit third-party reimbursement for their services. Starr (1982) wrote of a survey of 9,000 families conducted between 1928 and 1931, which found that nonphysician providers treated only 5.1 percent of all attended cases of illness. He concluded that "physicians had medical practice pretty much to themselves" (p. 127).

Kessel's argument is historical and written over 50 years ago, yet many analysts point to anomalies in recent medical school data as continuing evidence of control of entry by the medical profession. The large excess demand for medical school slots by qualified applicants to medical schools has been used to support the claims. Over time, large fractions of medical school applicants in the United States have been rejected; that is, there appears to be a substantial excess demand for medical school slots. Is this evidence of monopolizing control by the medical profession? Kessel's view certainly suggests that it is.

However, this simple story of professional control can be challenged. Hall and Lindsay (1980) argue that medical schools do not take larger proportions of applicants and medical school enrollments respond only partially to applicant demand because the administrators of medical schools are responding rationally to their economic incentives. As we have seen, medical school revenues come not so much from tuition paid by students, but from "donors"—that is, sources such as government agencies, alumni, businesses, and research organizations. For the most part, these donors are the true demanders of the output of medical schools—trained physicians. The donors may be especially interested in applicants from certain racial or ethnic groups, females, those with specific specialization interests, and those who indicate a willingness to return to shortage areas.

In 2005, Florida State University became the first fully accredited medical school established in almost 20 years. The following year, the Association of American Medical Colleges (AAMC) issued a policy statement calling for a 30 percent increase in student enrollments. In this new era in medical education, 15 medical schools started accepting students since the AAMC statement, in a clear response to market forces especially of perceived shortages in primary care and in underserved parts of the country. The additional demand for care from the ACA adds further support to the AAMC position. Together with expansion by existing schools, first-year medical school enrollment is projected to reach 21,000 by the 2017–2018 school-year compared to 16,700 in 2000–2001. First-year osteopathic medical school enrollment is expected to increase even more rapidly to almost 7,000 in 2017–2018 compared to just 2,900 in 2000–2001.

In summary, organized medicine historically exerted considerable influence over the supply of trained physicians. Such influence is consistent with a view of a profession seeking above-normal returns by trying to control entry of new physicians. However, data in recent years indicate that medical school enrollments respond to market forces. These data further suggest that it is no longer plausible to view medical education as controlled by a monolithic or conspiratorial medical profession.

Licensure and Monopoly Rents

Although licensure is not unique to the health care professions, licensure of physicians has received unparalleled attention. Starr (1982) provides a fascinating history of licensure legislation. The first licensure requirements for prospective doctors (though they had little effect and were minimally enforced) were passed in New York City in 1760. Subsequently, many states introduced licensing, often through state medical societies. After the 1820s, however, many of the same states modified or abolished licensure. It was not until after the founding of the AMA in 1847 and the last decades of the nineteenth century that stronger licensure laws were widely promulgated.

The primary controversies with respect to licensure relate to its role in limiting competition and the role of professional societies on state licensure boards. The conventional view held by many economists is that organized medicine has used control of licensure for self-interest by limiting entry (and by influencing the licensure requirements of potential competitor providers to the advantage of physicians). Some, however, have advanced a public interest argument for licensure—that is, as a result of information imperfections, the public demands quality controls. Licensure and certification help fill these information gaps.

Many economists believe that licensure and professional control over medical education ensure that physicians earn economic rents, which are payments to factors over and above those necessary to induce them to provide their services. These views were heavily influenced by the early work of Friedman and Kuznets (1945) and Kessel (1958). Friedman and Kuznets examined the relative return of physicians and dentists. After adjusting for training differentials, they estimated that about half of the 33 percent excess earnings of physicians between 1928 and 1934 represented economic rents.

Much has changed in the health industry since these earlier articles. However, the broader issues of monopoly rents and motivation for licensure remain controversial. Despite other empirical estimates supporting the rent hypotheses, several critiques of these studies have

appeared. Leffler (1978) argued that many earlier studies failed to take into account some important economic considerations that tend to reduce estimates of the return. These adjustments include the high number of hours worked by physicians, their expected mortality rates, and the progressive income tax structure (which took an increasing share of their incremental incomes).

A subsequent study undertaken by Burstein and Cromwell (1985) compared the internal rates of return of physicians to dentists and lawyers. The authors incorporate many adjustments into their estimates, including length of physician training, length of working life, and the earnings of medical residents. The rates of return were high compared to lawyers; for example, 12.1 percent versus 7.2 percent in 1980, the last year included in the study. The returns were high also for specialization based on board-certification requirements. This was true despite the rapid growth in physician supplies and the constraints imposed by third-party payers to contain costs over the study period. The authors concluded that "the conventional picture of medicine as a financially attractive profession is strongly confirmed" (p. 76).

This strong conclusion is further supported by a more direct test of physician pricing. Seldon and colleagues (1998) examined physicians' price–cost margins, defined as $(P - MC)/P$ where P represents price and MC represents costs. Under highly competitive conditions and marginal cost pricing (i.e., where P approaches MC), the price–cost margin is zero. If physicians have monopoly power and the ability to maintain price above marginal cost, the margin will be positive. The researchers estimated the margin at 23 percent overall (and from 13 to 54 percent across the nine regional markets in the study). These estimates indicate "nontrivial" levels of monopoly power that produced a welfare loss (due to insufficient care) to the U.S. economy of about $8 billion in 1996 dollars.

PUBLIC INTEREST AND SELF-INTEREST THEORIES OF REGULATION The specific issue of licensure is but a part of the broader issue of regulation reflecting the two competing theories: public interest versus self-interest motives. The public interest motive is based on theories of market failure, such as information failure. According to the public interest view, the demand for regulatory measures, such as licensure, stems from the limited information patients have about quality and the relatively high costs of obtaining information. Drawing on Akerlof's Lemons Principle introduced in Chapter 10, Leffler (1978) argued that asymmetric information will lower quality. Thus, a "state-enforced minimum quality standard is claimed to be an efficient response to costly quality information" (p. 173).

In contrast, the self-interest motives for licensure and other forms of regulation to reduce competition have long been accepted in economics but only relatively recently have been formalized (Stigler, 1971; Peltzman, 1976). This theory, discussed in more detail in Chapter 19, sees regulation as a return to special interests that provide financial and political support in return for favored legislation. Thus, a demand for political favors arises from the rent-seeking behavior of special interest groups. The effort and amount of resources expended by a special interest group are limited by the rents that would accrue from the favored legislation.

EVIDENCE OF PUBLIC INTERESTS VERSUS SELF-INTERESTS Paul (1984) tested the public interest versus self-interest theories using data on the initial decisions by states to license physicians, and he rejected the public interest theory. His findings show a strong negative association between the year of initial licensure and the number of AMA-associated physicians in a state per capita.

Graddy (1991) also tested the competing hypotheses by estimating the probability (and type) of regulation by states of six health care professions. Variables representing the

public-interest view included the profession's educational requirements in years (to capture the complexity of the service) and professional liability insurance rates (to represent the potential risk to consumers in the absence of regulation). The legislative environment was represented by variables such as the strength of the majority party and the rates of turnover of legislators. Variables such as the size of the occupation and its proportion of members belonging to a professional association represented private interests.

The statistical approach was possible because licensure practices for many professions vary substantially across states. Graddy found important roles for each of the categories of explanatory variables in determining the kinds of regulation. She found a higher probability of a stricter form of regulation as the profession's educational requirements are higher—a finding consistent with a public-interest motive. The overriding conclusion, though, is that no single dominant motive can be found for regulation. Legislators respond to organized interests, the public interest, and their own legislative environments.

Licensure and Quality

The Graddy evidence supports, in part, a public demand for regulation. We also know that physician board certification, or even board eligibility, increases remuneration, meaning consumers are willing to pay more for those with additional training and credentials. For example, O'Halloran and Bashaw (2006) simultaneously estimated the decision to become board certified and the returns to board certification. The likelihood of "investing" in certification is greater for physicians who stand to gain the most, including those who practice in more competitive markets (in order to stand out) and those with lower explicit and implicit costs associated with certification. Minority physicians earn a smaller reward from certification and they are less likely to become board certified. Overall, the authors conclude (p. 641) that "physician decisions to become board certified in their respective specialties rigidly follows a pattern consistent with human capital theory."

Does licensure actually improve the quality of care? Gaumer's (1984) review of the empirical evidence questions whether the goals of protecting the public and ensuring minimal standards of competency are being achieved. He found that (1) in spite of licensure, a substantial amount of deficient care occurs; (2) quality of care would not be impaired if the scope of practice of secondary (nonphysician/dentist) providers were increased; (3) the licensing process may "not accurately assess the practice competence of applicants" (p. 397); and (4) fees and provider incomes are higher in states with more restrictive licensure requirements (supporting the self-interests motive for regulation).

More specifically, with respect to the quality of physician care, he cites studies indicating that 5 percent of physicians are "unfit to practice," 8 to 22 percent of obstetrics patients and 61 to 65 percent of well-care patients received deficient care, and that 7.5 percent of all cases in two hospitals indicated physician-inflicted injury (p. 395).

Brennan and colleagues (1991) provide additional evidence on the quality of medical care in hospitals. Licensure is just one of many regulatory requirements intended to ensure quality. From a large number of randomly selected admissions, the researchers found that nearly 4 percent produced "adverse events," defined as injuries caused by medical management. Nearly 14 percent of these injuries led to death. The authors concluded that "there is a substantial amount of injury to patients from medical management, and many injuries are the result of substandard care" (p. 370).

Though no one is suggesting that eliminating licensure and other requirements will reduce such negative outcomes, regulation clearly does not ensure quality care. The national concerns with health care quality, as described in Chapters 10 and 14, and the policies promoted

by Congress to improve quality, are clear examples of the limits of regulation. Nevertheless, one could still ask whether tighter regulation would help. From his review, Gaumer concludes:

> Research evidence does not inspire confidence that wide-ranging systems for regulating health professionals have served the public interest. Though researchers have not been able to observe the consequences of a totally unregulated environment, observation of incremental variations in regulatory practices generally supports the view that tighter controls do not lead to improvements in quality of service.
>
> (p. 406)

As a result of the questionable effects of licensure on quality, changes in the health care environment, and the anticompetitive effects of restrictions on entry and restrictions on the scope of practice of potential competitors (e.g., podiatrists, nurse practitioners), the benefits of licensure are being re-examined. Svorny (1992) suggests that the benefits have been weakened by, among other things, the added liability that courts have placed on hospitals and HMOs for the negligent conduct of independent physicians and by the increased use of salaried physicians. In a stronger attack, Safriet (1994) concludes:

> Clearly these barriers serve no useful purpose, and in fact contribute to our health care problems by preventing the full deployment of competent and cost-effective providers who can meet the needs of a substantial number of consumers.
>
> (p. 315)

Have these conclusions changed over the past two decades—a period in which many states have expanded the scope of practice of nonphysician clinicians and in which the influence of organized medicine has been thought to wane? Not according to Svorny (2008) who describes the politics of licensure and the "turf wars" between medical physicians and other providers, writing:

> Medical licensure fails to meet expectations in the area of discipline and consumer protection. State medical boards' disciplinary efforts can arguably be said to protect clinicians more than consumers.
>
> (p. 11)

Other Physician Labor Issues

The prominence of physicians and their dominating role in treatment decisions have led to important research on a variety of labor issues. Many are associated with physician earnings, and we examine three of these issues below.

Specialization

Studies of physician specialty selection are especially important because of widespread beliefs that quality health care requires access to an appropriate mix of specialists. Policy effort in recent years has also sought to encourage more physicians to go into primary care, especially

in underserved areas. Some medical schools have responded to the challenge by favoring applicants who are committed to primary care.

The primary care challenge, however, is formidable due to the wide gap in earnings between specialists and generalists. For example, a Medscape survey (Pekham, 2015) indicates that medical specialists earned $284,000 in 2014 compared to $195,000 for primary care physicians. Estimates of lifetime earnings by Leigh et al. (2012) indicate that some specialties, e.g., neurological surgery, medical oncology, and radiation oncology, earn nearly $3 million more than those in family practice.

Economists recognize the role that nonpecuniary rewards, such as status and social responsibility, may play in decisions to specialize. Nevertheless, the economic focus is usually on the degree to which physicians respond to financial incentives. Decisions to specialize normally occur early in the physician's education and training career, so Nicholson (2005) was curious about medical students and their knowledge of physician incomes. He examined surveys of first- and fourth-year students conducted annually by large medical schools and found systematic biases in the responses. The students overestimated incomes in the 1970s but underestimated incomes by about 25 percent in more recent years. The study also showed a significant learning pattern with estimates that were much more accurate for fourth-year students than for those in their first year. The estimates were also more accurate for a specialty that a student was more likely to select.

How strong is the response to earnings potential, especially in light of the wide income variations by specialty? Economic theory suggests that a rational decision ought to be based on expected lifetime income, not simply the current earnings within a specialty, and several studies indicate substantial responsiveness. Estimated elasticities of entry into specialties with respect to changes in expected lifetime earnings are usually greater than one, leaving little doubt that physicians respond to income when making their specialty decisions. Nevertheless, one study of specialty choice provides some unique results on differences in specialty earnings.

Bhattacharya (2005) describes four possible explanations for the wide income disparities across specializations: (1) differences in hours worked, (2) differences in length of residency and other required training, (3) difference in the attributes and skills needed to perform in a specialty, and (4) barriers to entry into some specialties. The first three possibilities, which reflect competitive labor market forces, account for only one-half the observed differences in lifetime earnings. He concluded that the remainder of the differences results from differential entry barriers, suggesting that some specialist fees are excessive relative to competitive fees. From a policy perspective, a strong case can be made to pressure those specialty boards that impose relatively high entry barriers to increase their number of residency slots.

Private Practice or Employed

Many would be surprised to find that the majority of physicians, 62 percent, are employed rather than in self-employed private practice.[3] Hospitals have been aggressively purchasing physician practices to take advantage of incentives under the Affordable Care Act to form networks, such as the Accountable Care Organizations described in Chapter 12.

From the physician's perspective, the movement away from fee-for-service toward reimbursement methods that are tied to performance and accountability creates considerable

uncertainty. Those in private practice also have limited ability to deal with aggressive third-party payers that are constraining and even reducing payment rates. Other attractive features of salaried employment include a more regular work schedule and coverage for malpractice insurance. Of course, there is a trade-off in terms of independence and incomes. In 2014, private-practice primary care physicians earned 12 percent more and specialists earned 28 percent more than their employed counterparts.

It remains to be determined whether this transition away from private practice will actually improve coordination of care and lead to better care at lower costs. There is, however, one perverse consequence. Through facility fees, Medicare reimburses physician services provided by health systems at rates higher than those paid to independent practitioners even if the care site is the one previously used by the private-practice physician. As of early 2016, attempts to correct this anomaly have not yet been implemented.

Physician Income by Gender—The Increasing Role of Women

Women now account for about one-half of new medical school graduates. They also represented 30 percent of professionally active physicians in 2010 compared to just 11 percent in 1980. This dramatic shift toward gender balance will likely continue, and sociologists have eagerly studied a wide range of phenomena including the female physician's approach to patient care relative to the male physician. Economists have focused more narrowly on labor market issues, such as differences in earnings, job status, and hours worked.

Female physicians earn considerably less than male physicians. The Medscape survey indicated that male compensation in primary care was 32 percent higher than female compensation in 2014. Among the usual reasons for the gender gap is that women are more likely to choose the lower-paying specialties, and to work fewer hours than male physicians due to disproportionate burdens in raising a family. Sasser (2005) describes several mechanisms through which family responsibilities affect the gender gap. These include the fewer number of years in active practice that female physicians may anticipate and, thus, a reduced willingness by the female physician or her employer to invest in human capital. Greater household responsibilities could affect specialty selected and characteristics of the job environment sought. After controlling for specialty, practice setting, and demographic and professional characteristics, Sasser finds that female physicians "earn 11 percent less for being married, plus 14 percent less for having one child, and 22 percent less for having more than one child." The main determinant of these earnings differences are personal choices to reduce working hours sharply upon marriage and having children.

Despite Sasser's strong conclusions, the story of the gender earnings gap remains incomplete and complex (see Box 16.3). Timothy Hoff (2004) provides a rich example using hospitalists, a relatively new specialty consisting of hospital-based general physicians who focus on the care of hospitalized patients. He found that female hospitalists earn significantly less per year than their male counterparts even after controlling for demographic, professional, and job-related characteristics. Hoff also controlled for marital status and children. Furthermore, male and female hospitalists worked similar schedules and had similar levels of commitment. Thus, he concluded that the pay gap, at least in this new specialization, is real and not due mainly to personal and career choices. Hoff urges the medical establishment and policymakers to take pay inequality seriously and to develop mechanisms to address the problem.

BOX 16.3

The $16,819 Unexplained Gender Income Gap

Lo Sasso and colleagues (2011) analyze earnings data for physicians leaving residency programs in the state of New York over the ten-year period 1999–2008. The survey data provide information on many observable factors that might influence earnings, including specialty, hours worked, age, practice type, and employer location. By examining starting pay, the investigators avoid the difficult problem of controlling for some determinants of compensation, e.g., productivity, that are known over a period of time only after the initial hire.

Over the period covered by the study, an increasing proportion of females entered specialized fields. Nevertheless, the unadjusted female-to-male earnings ratio dropped from 87.4 percent in 1999 ($151,600 for females vs. $173,400 for males) to just 83.1 percent in 2008 ($174,000 vs. $209,300). After ordinary least squares regression was used to control for the observable factors, the unexplained pay gap of $35,400 in 2008 was reduced to $16,819. This unexplained gap is substantial especially as compared to the statistically insignificant differential estimated for 1999.

The investigators consider a variety of explanations including the possibility that unobservable factors account for the widening adjusted pay gap. One of the unobservable factors considered is the possible change in employment practices resulting from the influx of a large number of female graduates. In particular, the authors speculate (p. 198) that employment practices "may now be offering greater flexibility and family-friendly attributes that are more appealing to female practitioners but that come at the price of commensurately lower pay."

Conclusions

In this chapter, we used basic economic tools to provide important insights into a variety of health care labor issues, including the demand and supply of labor, optimal input decisions and factor substitution, and labor shortages. We examined two earnings issues as they related to specialization and the gender gap. The chapter has also addressed several aspects of medical education. In particular, we have tried to examine whether various characteristics of physician training and licensure are designed to increase barriers to entry into the profession, producing higher-than-normal rates of return.

We caution that rapid restructuring of the U.S. health care system, especially through managed care and post-managed care initiatives, is creating profound changes. As previously noted, physicians are increasingly organized in or affiliated with large groups that compete for managed care contracts.

At the same time, widespread purchases of physician practices by hospitals, with their reliance on salaried physicians, are other examples of change. Pay-for-performance, growth of high-deductible health plans, the emergence of Accountable Care Organizations, and the Affordable Care Act are other major developments that could have dramatic effects on health care delivery. As a result of these changes the economic power and professional influence of physicians have undoubtedly been affected in ways that are still evolving and

largely unknown. Clearly, this unprecedented pace of change represents new challenges for the health economist.

Summary

1 In 2014, 18 million people, representing 13 percent of total non-farm employment, worked in the health services industries.
2 The demand for labor is precisely the marginal revenue product of labor curve. It is closely related to the production function and is directly related to the price of the output.
3 The supply of labor tends to slope upward, implying that the higher the wage rate is in a given market, the more laborers will be forthcoming. Workers currently employed may choose to work more hours if higher wages are offered; other workers may be attracted from elsewhere by the higher wages.
4 There are basically two types of shortages: need shortages and economic shortages. Need shortages use a nonmarket, or noneconomic, definition of shortage.
5 One definition of an economic shortage is the excess of quantity demanded over the quantity supplied at the market wage rate. Stickiness in wages helps explain why the wage does not rise to equilibrium, thus automatically eliminating the shortage.
6 Meaningful shortages of professionals may exist even when supply and demand are in short-term equilibrium. If demand for a category of health professional expands over time and supply is slow to respond, the result may be a wage increase that is large relative to wage gains of other professionals.
7 Under labor monopsony conditions, a firm may report unfilled, budgeted positions, for example, for registered nurses, even when the firm is in equilibrium. A monopsony firm may announce that it wishes to hire more nurses even though it is unlikely to take the necessary steps to do so.
8 Medical education is heavily subsidized. Tuition is a relatively small source of revenues for medical schools; thus, the student pays only a small portion of the true costs of the investment in education.
9 Medical education is a good example of joint production. Medical schools produce medical education, patient care, and research.
10 Kessel argued that monopoly power was attained by organized medicine through licensure of physicians and control of access to medical education.
11 According to an alternative view of medical education, the donor-preference hypothesis, medical school revenues come not so much from tuition paid by students but from donors, such as government agencies, alumni, businesses, and research organizations. For the most part, it is these donors who demand and control the output of medical schools: trained physicians.
12 Licensure is a prominent example of the controversy of self-interest versus public interest views of regulation.
13 It is generally believed that licensure has given physicians economic rents. Licensure has not led to obvious improvements in quality.
14 There are wide differences in the earnings across physician specialties. Usual labor market explanations account for only one-half the variations, suggesting differences in barriers to entry among specialties.
15 There are also substantial gender differences in earnings. It is not yet clear whether these differences can be fully explained by the personal and professional decisions made by female physicians.

Discussion Questions

1 Give examples of ways in which labor and capital can be substituted for each other in the production of health services.
2 In the text, we considered only forgone income and tuition as costs of going to medical school. Enumerate other monetary and nonmonetary opportunity costs. Why are these opportunity costs relevant?
3 Why might demand for nurse labor by hospitals or other organized health providers be monopsonistic?
4 What is the marginal product of an input? Marginal revenue product? Why does the demand for a factor correspond to the marginal revenue product curve? What will determine whether the demand for a factor will be elastic or inelastic?
5 Why will a profit-maximizing physician firm want to equalize the marginal product per dollar spent across all inputs?
6 What is meant by the term *barriers to entry*? What are some entry barriers for someone who wants to be an obstetrician? For someone who wants to be a nursing assistant?
7 If barriers to entry into a profession were absolute so that entry would not be possible, what would the supply curve look like? What would the supply curve look like if entry into an occupation were free and easy? Thus, what role do barriers to entry play in explaining relative rates of return to an occupation?
8 Define *monopsony* and *marginal labor cost*. Why is the marginal labor cost in the case of monopsony above the supply (average labor cost) curve? What is the nature of the inefficiency or misallocation associated with monopsony power? Is there any inefficiency when the supply curve facing the monopsonist is perfectly elastic?
9 If there were no subsidies for medical education, would enrollments be larger or smaller? Would the return to medical education be larger or smaller? If physician education was not subsidized, would the economically warranted supply of physicians tend to emerge?
10 What are the social benefits and costs behind regulating the number of medical schools?
11 What is joint production? What does the term *joint production costs* mean? Given that medical schools engage in joint production of education, patient care, and research, what inferences can be drawn about the economies of scope in producing these three outputs?
12 In contrast to medical education, numerous night and part-time law schools have been established. Compare and contrast the various aspects of training that have led to these different educational systems.
13 What are some factors that help explain earnings differences across specialties? Why might the earnings differences persist over long periods of time?
14 The rate of return on investment in medical education exceeds that for other professions. What are arguments for and against government subsidies?
15 Female physicians earn considerably less than their male counterparts. Discuss some of the reasons that account for the differences. What kind of evidence would lead one to conclude that at least some of the difference is due to bias or discrimination?

Exercises

1 Consider the firm's demand (*MRP*) for labor, such as in Figure 16.2. If the demand elasticity is –0.5, what will be the effect of increased wages on total labor earnings?
2 Using Figure 16.3, graph and analyze the impact of an increase in the price of lab tests on the labor market.

3 Consider the market for highly skilled laboratory technicians. Graph the impacts on market wages if limitations on immigration were lifted. Would more or fewer services be provided? What would happen to the price?
4 In this chapter, we discuss how physicians' marginal products rise up to 25 hours and then slowly fall to zero at 110 hours. Graph both marginal and total products from this statement.
5 Using supply-and-demand analysis, model the equilibrium level of physicians' wages. What would be the impact on physicians' wages of more stringent policies on the employment of foreign medical school graduates?
6 Suppose that a medical school provides three outputs—patient care, education, and research—and that the total cost of the school is $100 million per year. If the school produced only education, its costs would be $60 million. If the school produced only patient care, its costs would be $30 million. If it produced only research, the costs would be $20 million. Joint costs for each pair would be $10 million.
   (a) What are the pure costs of education, patient care, and research?
   (b) What are the joint costs?
7 Suppose that the licensure requirements for health care providers were eliminated. Use supply-and-demand analysis to predict what may happen to the price and quantity of health care services. Are there other considerations—in particular, mechanisms—that could evolve to replace licensure?

Notes

1 There were also about 78,000 licensed doctors of osteopathy (DOs) in 2012 of whom 69,000 were active. Although we often combine DOs with MDs when considering physician supply and access to physician care, there have been major differences in the historical development and organization of the two groups. Unless otherwise indicated, our discussion will focus specifically on MDs.
2 Duffy (2011) notes that some of the deficient schools were deemed reparable, but the worst of the schools were "of such poor quality that closure was indicated."
3 Sources in this section include Peckham (2015); Elizabeth Rosenthal, "Apprehensive, Many Doctors Shift to Jobs with Salaries," *New York Times*, February 13, 2014 (http://nyti.ms/1evZjlo, accessed November 2016); and Margot Sanger Katz, "When Hospitals Buy Doctors' Offices, and Patient Fees Soar," *New York Times*, February 6, 2015 (http:nyti.ms/1zeXsM9, accessed November 2016).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 17

The Pharmaceutical Industry

In this chapter

- Structure and Regulation
- The Production of Health and Substitutability
- Drug Pricing and Profits
- Research and Development (R&D) and Innovation
- Cost Containment
- New Drugs and Health Care Spending
- The ACA and the Pharmaceutical Industry
- Conclusions

Prescription drugs and the pharmaceutical industry occupy increasingly important places in the health economy. Drug therapies traditionally have supplemented nutrition, sanitation, and medical care as methods for preserving health. Vaccinations for diseases such as smallpox represented early public health initiatives that saved thousands of lives. Insulin, developed in the 1920s, prevented the certain deaths that once accompanied diabetes; and a world without antibiotics, introduced in the 1940s, or the polio vaccination, in the 1950s, would be unthinkable.

Drugs are used to treat many diseases and conditions. Examples include chemotherapy for cancer, steroids for skin diseases, psychotropic drugs for mental health problems, beta-blockers for heart disease, clot busters for stroke, and protease inhibitors for AIDS. Some drugs prevent disease; some substitute for more invasive surgical procedures; some are used in conjunction with surgical and radiation treatments; while others provide treatment for conditions where no treatment was available previously. In recent years alone, observers view the introduction and widespread use of cholesterol absorption inhibitors to reduce the amount of cholesterol delivered to the liver for at-risk populations as a major breakthrough in the fight against coronary heart disease.

Despite these successes, the U.S. pharmaceutical industry has encountered intense media and legislative scrutiny. Pharmaceutical firms have been among the largest and most profitable businesses in the United States. As recently as 2001, the drug industry ranked first in various measures of profitability among *Fortune*'s industry groupings. Negative publicity, litigation problems, widespread efforts to contain drug spending, and loss of patent protection for several major drugs since then (see Box 17.1) are serious threats to profitability. Nevertheless, the 11 pharmaceutical firms among the *Fortune 500* in 2014 reported a median profit of 22 percent on revenues and 23 percent on stockholders' equity (*Fortune*, June 15, 2015, p. F-34). These were among the highest of all industries.

BOX 17.1

Patents and Media Attention

As we have noted, the pharmaceutical industry has a long history of superior financial performance. On account of perceptions of "exorbitant" prices and other questionable practices, drug companies are often the subject of unflattering media coverage. Here is one prominent example from early 2011.

In February 2011, the Food and Drug Administration (FDA) granted KV Pharmaceutical of St. Louis exclusive rights for seven years for the injectable form of a drug marketed as Makena. Makena reduces the risk of preterm delivery for expectant mothers with a previous premature delivery. Early delivery is a serious and growing problem in the United States so FDA approval sounded like welcome news.

However, Makena is chemically the same as another drug that was produced for years by a different firm and then withdrawn from the market. It was subsequently made by "compounding" pharmacies (those pharmacies that actually mix prescription lotions, creams, or doses for injection) at a cost of about $10 to $20 per shot (an expectant mother receives about 20 injections over the first 4–5 months of pregnancy). Some worry about the quality and consistency of the drug made by these pharmacies, but they also worry about the $1,500 price that KV initially established

for Makena. KV also warned specialty pharmacists that compounding the mixture would be unlawful. Following the media attention, KV announced a price cut to $690 per injection. The FDA also indicated in March 2011 that it will not take enforcement action against pharmacists that compound the drug.

Many important issues relate to patent protection (granted by the patent and trademark office) and exclusivity (granted by the FDA), and we will examine some in later sections of this chapter. At this time, we note that, despite its historical success, the pharmaceutical industry has recently faced unprecedented challenges to replace expiring patents with new revenue streams. Since 2011 alone, patents expired on a large number of blockbusters including Lipitor (to lower cholesterol), Advair (to prevent asthma symptoms), Zyprexa (to treat schizophrenia), Cymbalta (to treat anxiety and depression), Plavix (blood thinner), Abilify (to treat Schizophrenia), and Nexium (the "purple pill" for acid reflux disease).

Although 2015 was a banner year with 45 novel drugs approved by the FDA, the number of new drugs approvals has been running below historical rates and the FDA appears less likely to approve new products if there are significant side effects and if good therapies are currently available. As a result, the productivity of pharmaceutical investment in research and development (R&D) has decreased sharply. According to estimates prepared by the Deloitte Centre for Health Solutions (2014), the internal rate of return to R&D for 12 major firms declined from 10.1 percent in 2010 to 5.1 percent in 2014.

Sources: *Bloomberg Business Week*, "FDA Approves First Drug to Prevent Premature Births," February 6, 2011, businessweek.com/lifestyle/content/healthday/649631.html, accessed November 2016; Gardiner Harris, "Drugs' Cost and Safety Fuel a Fight," *New York Times*, April 4, 2011, nytimes.com/2011/04/05/health/05FDA.html, accessed November 2016.

Rapid growth in drug spending has made the pharmaceutical industry a convenient target for the budgetary challenges facing patients and insurers. The introduction of new and expensive drugs, increasingly promoted through direct marketing to consumers, has only added to the expenditure burden and heightened criticism of the industry.

This burden becomes especially severe for those who suffer disproportionately from chronic and other conditions that fuel the use of drugs. Policies to deal with these pressures include the 2006 expansion of Medicare (Part D) to include outpatient prescription drug benefits (see Chapter 20), and proposals to regulate prices as well as to permit the re-importation of drugs from Canada and other countries. Private insurance initiatives include higher patient copayments, increased emphasis on generic products, and new strategies, such as the development of drug formularies. Most managed care plans adopted formularies, that is, approved lists of drugs, by the late 1990s. Patients may have difficulty in obtaining reimbursement for any drugs that are not on the list.

Pharmacoeconomics, which includes cost-benefit, cost-effectiveness, and cost-utility analyses, plays an increasingly important role in pharmaceutical decisions, but policy must address other questions. This chapter selects several of the most general interest to health economists. After describing the structure and regulation of the pharmaceutical industry, we focus on the following areas:

1 The role of pharmaceutical products in the production of health, patient choices of drugs under various insurance schemes, and the effects of technological change on the use of drugs.
2 Drug pricing issues, including price discrimination by sellers and price regulation by the government.

3 Pharmaceutical research, the determinants of innovation, and the effects of price regulation on innovation.
4 Cost containment through use of generic products and other measures.

We conclude with recent evidence on the effects of drugs on health derived from international comparisons on drug utilization.

Structure and Regulation

In 2014, spending on prescription drugs amounted to $298 billion or 9.8 percent of national health expenditures (NHE). Although this share is up from 8.8 percent in 2000 and just 4.7 percent in 1980, it is down from a peak of 10.3 percent in 2006 despite a rapid increase in prescription drug spending of 12.2 percent in 2014. Consumers' out-of-pocket costs for drugs represented 18 percent of total spending on drugs in 2014, and accounted for 14 percent of all out-of-pocket costs.

Although prescription drug spending has stabilized at 9–10 percent of NHE in recent years, its high long-run growth rate relative to other categories of health spending created considerable interest in the sources of these increases. Analysts have found that most of the increases have been due to greater use of drugs and to new products rather than to higher prices of existing products. Nevertheless, rising expenditures and high out-of-pocket costs help create public perceptions that something is seriously wrong with the conduct of pharmaceutical firms, and that stiff measures are needed to contain drug costs (see Box 17.2).

BOX 17.2

Martin Shkreli and Valeant Pharmaceuticals

The September 20, 2015 *New York Times* article, "Drug Goes from $13.50 a Tablet to $750, Overnight," created a firestorm. Turing Pharmaceuticals led by Martin Shkreli acquired the drug Daraprim. Daraprim is the standard treatment for toxoplasmosis—a parasitic disease that could be life-threatening for those with compromised immune systems, e.g., with AIDS and certain forms of cancer. Daraprim is no longer patent protected and actually cost $1 per pill several years ago but, with only about 10,000 prescriptions per year, other pharmaceutical firms have not entered the market. Despite being called the "most hated man in America" by some media outlets, and subsequently arrested for alleged securities fraud, Mr. Shkreli told *Forbes* that he should have raised prices even higher because "my shareholders expect me to make the most profit . . . that's the ugly, dirty truth."

With the spotlight on Mr. Shkreli, the public was learning that the Daraprim episode is not unique. Valeant Pharmaceuticals International became the largest publicly traded company in Canada, reaching a market value of $90 (U.S.) billion in mid-2015. Valiant's growth was fueled by numerous acquisitions of pharmaceutical and medical supply companies. As part of its strategy, Valeant also bought the rights to older

generic drugs and subsequently raised prices by substantial amounts. For example, in 2015, Valeant raised the prices of two heart medications (Isuprel and Nitropress) by 525 percent and 212 percent after purchasing the rights to these drugs from another firm. The fallout from the negative press and Congressional hearings into its business practices reduced Valeant's market value by 70 percent by October 2015.

Sources: Andrew Pollack, "Drug Goes from $13.50 a Tablet to $750, Overnight," *New York Times*, September 20, 2015, http:nyti.ms/1V3cJvC, accessed November 2016; Dan Diamond, "Martin Shkreli Admits he Messed Up: He Should've Raised Prices Even Higher," *Forbes*, December 3, 2015, forbes.com/sites/dandiamond/2015/12/03/what-martin, accessed November 2016; Carly Helfand, "Shkreli-Shaming Spills over onto Valeant as Dems Call CEO to Account for Price Hikes," *FiercePharma*, September 28, 2015, fiercepharma.com/story/shkreli-shaming-spills-over-valeant, accessed November 2016.

With its long history of relatively high profits and rich set of features—patent protection, high research and development spending, intense product promotion, and heavy regulation—the pharmaceutical industry always has drawn the attention of economists in the field of industrial organization. Scholars describe levels of competition in an industry; how the competitive environment influences decisions on prices and other decision variables, such as advertising, research and development (R&D), and quality; and the consequences of these decisions for socially efficient allocations of resources.

Competition

The level of competition often holds the key to firm and industry behavior. To measure competition, economists need to look at meaningful industry groups. Prior to 1997, the Standard Industrial Classification (SIC) served as the standard. The Department of Commerce has since replaced the SIC codes with the North American Industry Classification System (NAICS). The NAICS codes range from two to six digits, with each successive digit representing a finer degree of classification. The "pharmaceutical preparations" industry, NAICS Code 325412, formally consists of "establishments primarily engaged in manufacturing in-vivo diagnostic substances and pharmaceutical preparations (except biological) intended for internal and external consumption in dose forms, such as ampoules, tablets, capsules, vials, ointments, powders, solutions, and suspensions."

The four-firm (C_4) and eight-firm (C_8) concentration ratios for any selected six-digit NAICS codes indicate the share of industry output produced by the four or eight largest firms. Analysts use these concentration ratios, shown in Table 17.1, for pharmaceuticals and several other six-digit industries, to gauge competition. The pharmaceutical industry is one of the largest manufacturing industries. As measured by concentration ratios, it also appears to be much more competitive than many others.

Another method used to measure competition is the Herfindahl-Hirschman Index (HHI). Compare an industry with four firms, each with 25 percent of the market, to a second industry, also with four firms, but where one firm has 85 percent of the market and the other three each have 5 percent. Both industries have a four-firm concentration ratio of 100 percent. However, one might guess that the one in which the leading firm has 85 percent of the market is more monopolized. The HHI incorporates differences in the size distribution of firms by squaring the market shares of each and adding them together, so that the lowest value approaches 0 (thousands of tiny firms) and the highest value approaches 10,000 (with a pure monopoly). In the previous example, the respective HHIs are 2,500 (the four equal sized firms) and 7,300 (the very large firm, with three smaller ones).

Table 17.1 Concentration in Selected Manufacturing Industries: 2002

NAICS Code	*Industry*	C_4	C_8	*HHI*	*N*	*Shipments (in $ billions)*
325412	Pharmaceutical preparation mfg.	36	53	530	731	114.7
311230	Breakfast cereal mfg.	78	91	2,521	45	9.1
324110	Petroleum refineries	41	64	640	88	193.5
334111	Electronic computer mfg.	76	89	2,662	934	32.3
334220	Radio & TV broadcasting & wireless	43	55	584	823	32.1
325510	Paint and coating mfg.	37	55	505	1,149	19.9
325611	Soap and detergent mfg.	61	72	2,006	699	16.6
336111	Automobile mfg.	76	94	1,910	164	88.1
336112	Light truck & utility vehicle mfg.	96	100	W	69	137.1
336411	Aircraft mfg.	81	94	W	184	64.3

Note: W = withheld to avoid disclosure of individual firm data. The undisclosed HHI value will undoubtedly be very high.
Source: U.S. Bureau of the Census, 2002 Economic Census, "Concentration Ratios: 2002," Report EC02-31SR-1 (May 2006).

Table 17.1 shows that compared to other well-known industries, in addition to relatively low concentration ratios, there are a relatively large number of firms (N) and the HHI (for up to the 50 largest firms) is relatively low. Do these data indicate substantial competition? Most analysts would argue that for pharmaceuticals they could be especially misleading. Drugs in different therapeutic categories usually are not substitutes for each other. Concentration ratios for narrower drug classes are better indicators. When such data are available, they still can show considerable competition. In many cases, however, the concentration ratios will be higher, sometimes much higher. Schweitzer (1996) illustrates this with a class of drugs used to control hypertension. The top four firms controlled 91 percent of the market in 1992. Why? Patents and other barriers to entry often restrict competition.

Barriers to Entry

A barrier to entry is any factor that impedes the entry of new firms into an industry or product market. Patent protection granted by government represents a classic example. To gain further protection, pharmaceutical firms adopt a common business strategy of surrounding a product with patents on many variations of that product.

A patent forms a legal barrier. Advertising and promotion also can create economic barriers when they successfully increase brand loyalty. Pharmaceutical promotion differs from that of typical consumer goods because pharmaceutical firms direct much of their marketing

at physicians rather than patients, the end users through "detailers"—pharmaceutical representatives who directly visit physicians' offices. Critics of this practice believe that detailing may lead to questionable financial arrangements that encourage the physician to prescribe a particular product, possibly in place of cheaper drug or nondrug substitutes.

Pharmaceutical firms also reach physicians by distributing samples, by direct mail, and by advertising in medical journals. Following the FDA's relaxation of rules governing advertisements through the media in 1997, the industry responded by increasing advertising in newspapers, on radio, on television, and even on freeway billboards, all aimed directly at patients (see Box 17.2). The medical community and other critics of direct-to-consumer (DTC) advertising have raised concerns about the misinformation, confusion, and unnecessary or even harmful treatment that could result from such advertising.

As a last example of protection from competition, the regulation of drugs itself can create entry barriers. The Food and Drug Administration (FDA) approval process for a new drug is costly and time consuming. A new firm will find it difficult to marshal the financial and expert resources needed to go through the process and especially to have a portfolio of products under development to spread risks. According to the industry, only "five in 5,000 compounds that enter preclinical testing make it to human testing" and only one of these five ultimately is approved as a drug (PhRMA 2010, p. 16). Fewer still ever become profitable. Not surprisingly, such long odds create formidable deterrence to new drug development, and new pharmaceutical firms often concentrate on generic products.

BOX 17.3

Direct-to-Consumer (DTC) Advertising

Prior to 1951, the distinction between over-the-counter and prescription drugs was not as well defined as it is today. The FDA did require that certain highly potent and potentially dangerous drugs be available only through prescription, but the decision for many others was left to the producer until the 1951 Durham Humphrey Amendment to the Food, Drug, and Cosmetic Act of 1938. In the following years, an increasing proportion of medications were available only through prescriptions. However, until the 1990s, the pharmaceutical industry overwhelmingly concentrated its promotion efforts on doctors, largely through "detail men" who would visit physician offices. This strategy reflected the prevailing view of the medical decision-making model as one based on the authority of the physician over a passive patient.

It was not until the 1990s that marketing managers began to re-evaluate the potential of DTC advertising. There were two important developments in this process: (1) the growth of managed care, which constrained consumer choices and put downward pressure on drug prices, and (2) the growth of consumerism in general, but especially in health care. Nevertheless, DTC marketing amounted to only $363 million in 1995, with just 15 percent directed to the broadcast media. The major impetus came in 1997 after the FDA made it easier for broadcast ads to meet requirements regarding a summary of the risks and benefits of the advertised product. For example, the ad could now direct consumers to a toll-free number or to a website for such information. DTC advertising, especially on television, grew rapidly, reaching $4.8 billion in 2006 for the research-based pharmaceutical firms (out of $12 billion spent on all marketing and promotional activities).

DTC television advertising remains one of the most controversial and visible practices of the pharmaceutical industry. It is allowed only in the United States and New Zealand among advanced countries. The drug industry maintains that DTC advertising "creates awareness of diseases and treatment options and empowers patients with information."

While even critics may agree that television ads can inform patients and reduce the stigma associated with some conditions (e.g., sexual dysfunction, incontinence, and mental disorders), they also charge that they pressure physicians to overprescribe or to prescribe expensive drugs when cheaper alternatives are available. Even worse, critics claim that DTC advertising manufactures diseases and creates a lifelong dependency on expensive drugs for conditions that were once considered normal or natural. A *New York Times* investigation of marketing practices for attention deficit hyperactivity disorder (ADHD) provides a powerful example.

After noting medical concerns about the rapid rise in ADHD diagnosis in children to those with minimal symptoms, the *Times* describes the marketing of psychostimulants to treat ADHD symptoms through various print and media channels as well as other practices. Both doctors and parents are the targets. Side effects are downplayed and the drugs are often marketed as "safe" or "harmless." But every major manufacturer of ADHD drug has been cited multiple times by the FDA for false and misleading advertising. The industry is now targeting adults, which according to the *Times* could be even more profitable than the children's market.

Sources: Donohue (2006), PhRMA (2008), Ventola (2011), and Alan Schwarz, "The Selling of Attention Deficit Disorder," *New York Times*, December 14, 2013: nytimes.com/201312/15/health/the-selling-of-attention, accessed December 15, 2013.

Regulation

The pharmaceutical industry is one of the most heavily regulated of all industries. Governments regulate most firms for worker safety and health concerns, but pharmaceutical products face further oversight by the FDA. Following a public scandal over adulterated food products and dangerous medicines with unknown contents, the federal government introduced the Food and Drug Act of 1906. The act did nothing to prevent the public from dangerous medicines. It did not even require formal testing but dealt mainly with labeling. Requirements for testing and safety were introduced with the Federal Food, Drug, and Cosmetic Act of 1938. However, these requirements were left mainly to the drug companies.

Two events accelerated regulatory change. Exposure of questionable drug industry practices in hearings held by Senator Estes Kefauver in 1959 was soon followed by the thalidomide tragedy. Thalidomide, a tranquilizer widely used in Europe to treat morning sickness in pregnancy, was discovered to cause severe defects in babies, who were sometimes born with deformed, flipper-like limbs. The drug was available on an experimental basis in the United States at the time. Fortunately for the United States, the number of thalidomide babies was relatively small. The FDA had delayed approval, and the distributor withdrew the product quickly after reports of the European experience.[1]

Although the thalidomide tragedy was averted in the United States, Congress nonetheless approved amendments in 1962 that gave the FDA increased control over the introduction of new products. The new legislation required much more testing and extended the FDA's authority to regulate premarket testing (including generic drugs). Equally important, the legislation for the first time required evidence of efficacy.[2]

FDA review has become a lengthy, complex process. Following the discovery stage during which new chemicals are synthesized, the firm conducts preclinical animal studies involving short-term toxicity and safety tests. The drug firm next must file an application with the FDA to conduct clinical trials. If approved, the trials are conducted in three phases. Phase I begins with small groups of healthy volunteers and focuses on safety and dosage. Phase II trials involve a larger number of subjects, often several hundred, who have the targeted condition, and concentrates on the drug's efficacy.[3] Phase III trials usually are conducted on thousands of patients in different settings so that safety and efficacy can be determined more precisely.

If these trials indicate safety and efficacy, and the drug's safety is supported by long-term animal studies, the company submits a New Drug Application (NDA) containing all the data and results to the FDA. The FDA review usually takes more than a year. Total development time for a new product stands at about 14 years, nearly double the eight-year period in the 1960s (DiMasi, 2001).

These requirements provoke considerable controversy and provide obvious trade-offs between the goals of protecting the consumer and rapid innovation. The economic approach is to weigh the gains in safety and efficacy against the cost of delaying patients from utilizing useful products. Economists also express concern about the potential stifling of innovation caused by regulation and its adverse effects on competition.

In a classic study of the 1962 amendments, Peltzman (1974) found a sharp decline in new product development, especially of innovative drugs, after 1962, as well as higher prices from the decreased competition. These consequences far outweighed the benefits of reduced spending on ineffective drugs, creating a net welfare loss of about 6 percent of total drug sales.

The FDA recognized these problems and in the mid-1970s developed policies to accelerate the review of "important" drugs. Dranove and Meltzer (1994) found that important drugs reach the market about three years sooner than other drugs. Thus, they argue that the losses resulting from delays in the approval process have been overestimated. A 1984 act also eliminated the full range of tests for generic products that were required by the 1962 amendments.

To expedite the review process, 1992 legislation and the Modernization Act of 1997 provide the FDA with additional resources derived from user fees levied on the industry. This has considerably reduced approval times. The Food and Drug Administration Amendment Act of 2007 included components that enhanced FDA authority and gave it significant increases in users' fees to conduct comprehensive reviews of drugs and medical devices. In 2012, as part of the Food and Drug Administration Safety and Innovation Act, the FDA introduced a "breakthrough therapy" designation to speed up the process for drugs targeting serious or life-threatening conditions.

Philipson et al. (2008) found a very favorable trade-off between approval times and safety of legislative changes between 1992 and 2002. More rapid access to drugs saved between 140,000 to 310,000 life-years compared to an upper bound of 56,000 life-years lost due to harmful effects of drugs before they were withdrawn from the market. However, Olson's (2008) work reminds us of the risk. She found that a reduction in review time of one standard deviation increased serious adverse drug reactions by 21 to 23 percent, and hospitalizations and deaths from these reactions by about 20 percent each.

The Production of Health and Substitutability

We have seen that spending on prescription drugs is increasing rapidly and that drug firms have some monopoly power. Before we examine the exercise of that power, we turn to the

role of prescription drugs in producing health and their relationship to other medical inputs using the concept of a health production function.

Recall that the patient's demand for health leads to a demand for health inputs such as drugs and medical care. Consider the following production function for a patient with chronic and severe low-back pain—one of the most common reasons for physician office visits and hospitalization:

$$HS = f(D, M) \tag{17.1}$$

where HS represents the individual's health status in the current period, D represents prescription drugs, and M represents all other medical inputs in this period, given existing technology and medical know-how. If no drugs or medical inputs are applied, the patient might experience considerable pain and be unable to perform many normal tasks, including work.

Assume that this patient's health can be improved by medical intervention. Suppose an individual consumes the amount of drugs, D_1, and the amount of other medical inputs, M_1, as noted at point E of Figure 17.1. How do the drugs and the medical inputs substitute for each other?

Figure 17.1 shows three different effects of drug products and their relationship to other medical inputs. Isoquant 1 shows that drugs (e.g., narcotic analgesics or muscle relaxants) must be used in a fixed proportion to other inputs (e.g., physical therapy, counseling, and surgery in some cases). Here, inputs D and M are perfect complements with no substitutability between them.

At the other extreme, isoquant 2 reflects a production function where inputs are perfect substitutes: The marginal rate of technical substitution (MRTS) is constant, meaning that drugs substitute for the other inputs at a constant rate. (Depending on the prices of each,

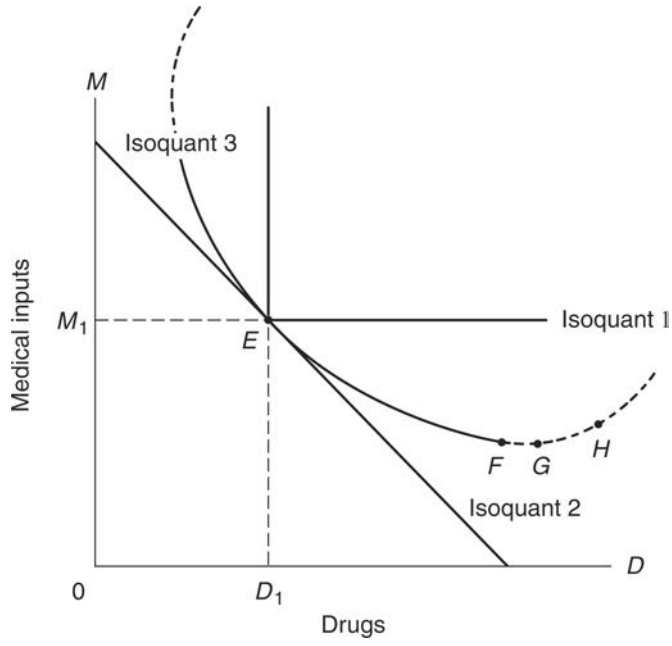


Figure 17.1 Substitution between Drugs and Other Medical Inputs

a patient would use one or the other, but not both.) Finally, the solid portion of isoquant 3 reflects an intermediate situation where *D* and *M* can substitute for each other, but where the MRTS is diminishing. Fewer and fewer amounts of *M* are needed to substitute for a unit increase in *D* as more drugs are used.

It seems unlikely that *D* and *M* are either perfect complements or perfect substitutes. Many conditions likely involve some substitutability, although the extent may vary widely among conditions and even patients. Pharmacological and other clinical studies must determine not only substitutability, but also the "uneconomic" portions of an isoquant, meaning those combinations that never should be selected.

Such combinations arise when additions of *D* over some range (e.g., dashed segment *FG*) will not benefit or harm the patient. Here, the marginal product of *D* is zero and the isoquant becomes horizontal. It is also possible that increases in *D* beyond some point may harm the patient and require more medical intervention to maintain the same health status. In this region (arc *GH*), the isoquant will become positively sloped. Similar logic may apply to the vertical portion of an isoquant with increases in *M*. Patients will not want to be in the regions shown by the dashed segments.

Least-Cost Production

How will the patients and their providers choose? To abstract from the many possible levels of health, as well as the inherent uncertainty of medical practice, assume that the patient/provider believes that it is reasonable to attain HS_1 in Figure 17.2. The rational patient seeks to find the combination of *D* and *M* on HS_1 that minimizes spending. Without insurance coverage for either *D* or *M*, the total cost (*C*) of care can be written as:

$$C = P_D D + P_M M$$

or

$$M = C / P_M - (P_D/P_M)D \tag{17.2}$$

where *D* and *M* are quantities of drugs and other inputs, and P_D and P_M are their respective prices. If, for example, P_D = \$50 and P_M = \$100, the slope of the budget line in equation (17.2) is –(50/100) = – 0.5.

The cost-minimizing combination is at *E*, where the isoquant, HS_1, is tangent to the budget line, *AB*. The optimal inputs are D_1 and M_1 and we calculate total spending by multiplying these quantities by their respective prices. At *E*, the numerical slope of the budget (P_D/P_M = 0.5) equals the MRTS, the slope of the isoquant. Suppose, for example, that D_1 = 4 and M_1 = 6, so total costs for HS_1 equal:

$$\text{Costs} = (\$50 \times 4) + (\$100 \times 6) = \$800$$

If prescription prices increase above \$50, the budget line will become steeper and the rational patient will try to substitute more medical care by moving to the left on the isoquant to a point such as *E′* (and vice versa to *E″* if medical care prices increase).

Insurance and Substitutability

Assume now that like most Americans the patient has insurance coverage. Begin with a policy that covers a constant proportion (e.g., 80 percent) of spending on either *D* or *M*. Out-of-pocket patient costs are \$10 for each prescription (20 percent of \$50) and \$20 for a medical

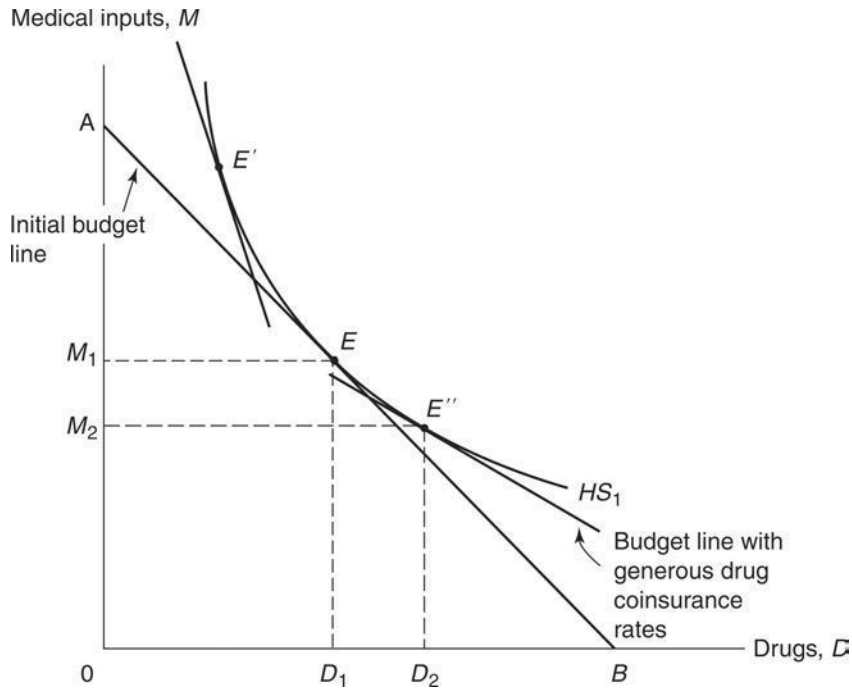


Figure 17.2 Cost Minimization

visit (20 percent of $100). Because the slope of the patient's budget line does not change, the optimal choice remains at point *E*, and the patient continues to buy four units of drugs and six visits. Total drug costs will continue to be minimized with the patient paying 20 percent ($160) and the insurer paying 80 percent ($640) of the $800 total bill.

However, *D* and *M* often are not treated uniformly under traditional health insurance. Consider a policy that pays 80 percent of medical costs but requires a deductible of only $5 (copayment) for each prescription. The patient's drug price is the $5 deductible regardless of the actual price of the medication. If the patient's out-of-pocket drug costs diminish, the numerical slope of the cost-minimizing budget line diminishes (in our example, it is now –5/20 = 0.25). The patient will have an incentive to substitute *D* for *M* at *E*″. Continuing with the example, let D_2 increase from 4 to 5, and M_2 fall from 6 to 5.75.

The patient's cost burden diminishes from $160 to $140 or:

(5.75 visits × $100/visit × 0.2 coinsurance rate) + (5 units of drugs × $5/unit)

However, the total cost of care (patient plus insurer) increases from $800 to $825. We know this is true because we already determined that *E* is the least costly combination to provide HS_1.

Similarly, if prescription prices (to the insurer) increase, say to $100, the patient still pays $5 and will remain at *E*″ with the insurer picking up the increased drug costs. Patients have no incentive to economize by making substitutions and moving toward *E*. The higher the prescription price, the greater is the distortion.

A similar distortion toward excessive levels of *M* and greater total costs occurs when the patient's coverage excludes or limits drug benefits. Here, the savings from reducing *M* will more than offset the additional drug spending from improved drug coverage.

Technological Change

At the turn of this twenty-first century, a new category of experimental drugs, blood vessel inhibitors, generated extraordinary excitement in the medical community by fighting both cancer and heart disease. Technology often is associated with major breakthroughs. More often, however, new drugs are similar to existing drugs, but they may produce somewhat better outcomes (if only for some patients) or reduced side effects.

With technological improvements, fewer inputs are needed to produce a given health outcome, or outcomes that were previously unattainable are now attainable. For example, begin with HS_1 in Figure 17.3 and let $HS_1{}^*$ represent all combinations of inputs with a new drug that leads to the same health status as HS_1. If the cost-minimizing ratio of inputs at a given price ratio remains unchanged, so that it lies along the ray $0E$ (denoted $(M/D)_0$), the innovation represents a neutral technological change. As drawn, the new drug saves a relatively high amount of the medical input at any given price ratio. That is, drug utilization increases relative to medical care as the patient moves to E^* and substitutes D for M. As noted in the figure, the lower ratio of M to D is reflected in the less steeply sloped ray, $(M/D)_1$.

New technology can increase costs for two reasons. First, it can routinely provide health levels that were unattainable previously (e.g., $HS_2{}^*$ at point E^{**}). It may require much more drug use and possibly increased use of M, as well. When health improvements are dramatic or when drugs treat serious conditions that were not treatable previously, cost concerns are likely to be far less troublesome than those leading to only marginal improvements in health.

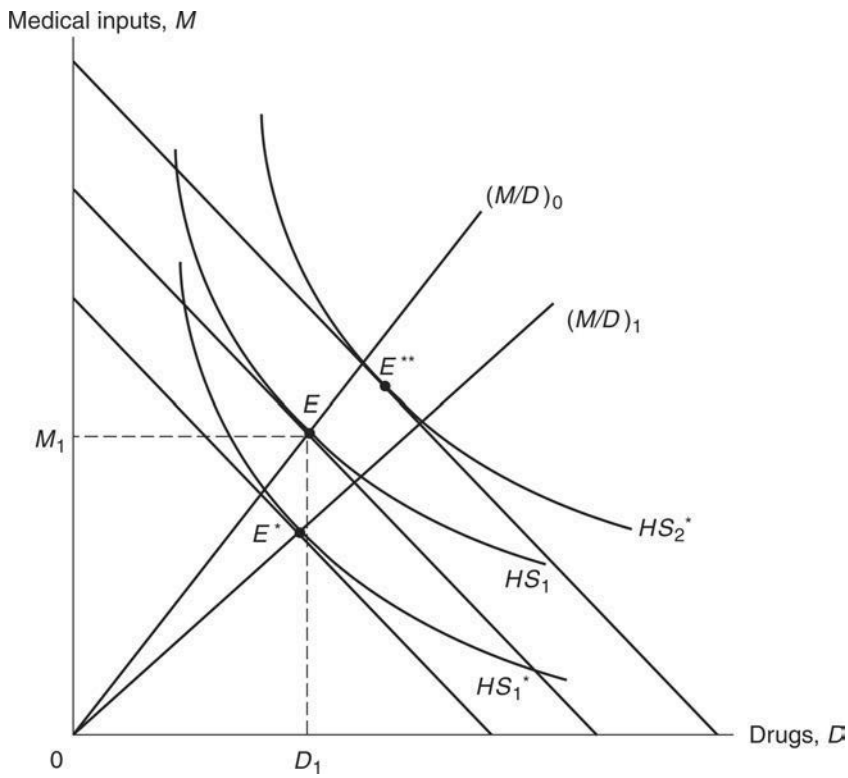


Figure 17.3 Technological Change

The second cost pressure comes from insurance. As we have seen, a patient with a constant copayment will not face any price increases of the new drug. Assume that the slope of the budget line in Figure 17.3 reflects the patient's copayment of $5 and the patient moves from E to E^*. If the price of the new drug is high enough, the total cost of care at E^* could be substantially higher than total costs at the original equilibrium at E.

For conventional goods, where the consumer pays the entire price out-of-pocket, such technological changes will not be introduced because they will not be demanded. With insurance, the determination and elimination of cost-inefficient technology are far more difficult. A drug-maker may market a socially cost-inefficient drug successfully simply because it is more convenient for the patient to reduce other services and take more medications at E^*.

Drug Pricing and Profits

Drug pricing and profitability undoubtedly generate the strongest reactions among the public and the media (see Boxes 17.1 and 17.2). News reports (e.g., "Doctors Denounce Cancer Drug Prices of $100,000 per Year," *New York Times*, April 25, 2013) paint pharmaceutical companies as exploiting patients through patents and other strategies that reduce competition. Many studies have found that pharmaceutical profits, as reported in financial statements, are consistently among the highest of all industries.

This relatively high return is often attributed to monopoly power, but the profit picture is far less clear. Conventional accounting methods treat R&D and advertising and promotion as current expenses to be "expensed" even though, like physical investment, they provide returns in future years. Expensing can be thought of as an extreme form of accelerated depreciation where all of the "R&D capital" is used up in one year. It raises rates of return by reducing taxes. When Clarkson (1996) made adjustments to capitalize and depreciate these "investments," the industry's return remained higher than average but well below the adjusted returns for the highest industries.

Others argue that if drug R&D is riskier than other types of investments, it requires a higher rate of return to attract capital into the industry. Without trying to sort through all the measurement nuances, it seems reasonable to conclude that pharmaceutical firms earn above normal rates of return but that their profitability has been exaggerated by simple accounting comparisons.

Monopoly Pricing

We begin with a firm selling a single product (or a composite of products) at a uniform price to all buyers. Figure 17.4 shows the demand and cost conditions facing the firm. Demand is a negatively sloped curve for several reasons. Even though patients with fixed copayments do not face higher out-of-pocket prices and will have a perfectly inelastic demand, others have more limited coverage or no drug coverage at all. Patients with limited (or no) coverage will likely substitute generic or over-the-counter products as a drug's price increases. Some may reduce utilization by not complying with the medication regimen. The drug supplier also must consider purchasing decisions by managed care organizations and other large buyers, such as hospitals, which can be sensitive to price changes.

On the cost side, the marginal cost of manufacturing and distributing the product is usually relatively low—about half the total cost. R&D and various promotion costs are substantial. In Figure 17.4, we show the marginal cost (MC) as constant and the average cost (AC),

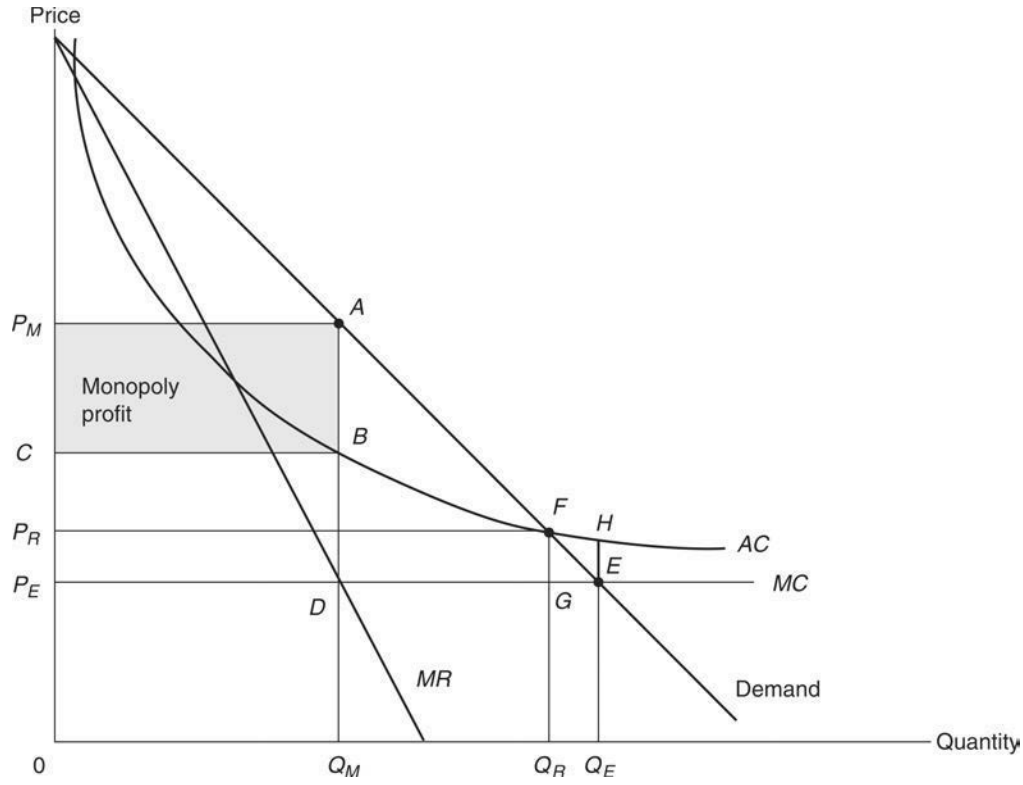


Figure 17.4 Drug Pricing

due to high fixed costs of R&D and promotion, as a downward-sloping curve. The profit-maximizing output occurs where *MC* equals *MR*, resulting in P_M and Q_M. Assuming that the drug supplier earns economic profits, the price must lie above average cost. The shaded rectangle $P_M ABC$ shows economic profits. Consistent with the hypothesized demand and cost structures, the gap between price and the low marginal cost will be large.

The profit-maximizing model also predicts that the difference between price and marginal cost varies inversely with the elasticity of demand. Lu and Comanor (1998) examined pricing decisions on new products, and their findings support profit-maximization. Initial (launch) prices are considerably higher for products that represent large, therapeutic gains than prices for new "me-too" drugs that are similar to available products. Why? Demand will be relatively inelastic for a product that provides significant benefits as compared to other products. Launch prices are also much higher when few branded substitutes exist. This factor similarly reduces a product's demand elasticity and requires a higher price for profit maximization.

Price Discrimination

A firm may be able to increase profits beyond the level described in Figure 17.4. One of the most interesting features of the pharmaceutical industry is third-degree price discrimination

(also known as market segmentation) where different groups of buyers are charged different prices. Before the enactment of Medicare Part D drug benefit legislation, the media routinely ran news stories about American seniors flocking to Mexico, where prices are much lower, to buy drugs. Hospitals or managed care groups are often charged less than retail pharmacies, and prices for drugs used in veterinary medicine can be much lower than prices for similar products packaged for human use.

What accounts for the wide variations in price? One explanation is straightforward. If a firm can distinguish between markets with different demand characteristics, and can also limit arbitrage (third-party resale at lower prices in higher-priced markets), it can increase profits by charging different prices. Assume, for simplicity, that the firm described in Figure 17.4 sells only in the United States and Mexico. Figure 17.5 separates the total demand into the U.S. and Mexican demands. With higher incomes and better insurance, the demand is relatively inelastic in the United States. Assume further that the marginal costs of production and distribution remain constant and are equal in both countries, and that prices are not regulated in either market.

Profit maximization occurs where *MR* equals *MC* in each market, resulting in quantities Q_{US} and Q_X. Even though marginal revenue will be equal in the United States and Mexico, the price is higher in the market with the less-elastic demand (United States).[4] Total profits must be greater than those obtained under uniform pricing.

Monopsony Pricing and Price Controls

Price discrimination is not the only possible explanation for price differentials. Prices in some foreign countries can be lower because their governments regulate prices or their national health plan serves as a monopsony buyer. Continuing with Figure 17.5, suppose that the Mexican government imposes price controls. Conceptually, it can drive price as low as the marginal cost, further increasing the price differential with the United States. Critics charge that by failing to control prices in a similar manner, consumers in the United States bear the burden of the development costs, and that the United States subsidizes other countries.

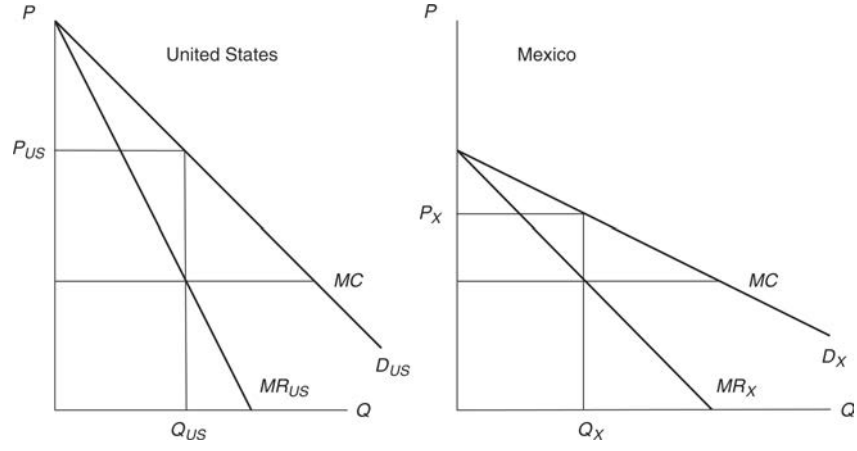


Figure 17.5 Price Discrimination

Proposals have been made to limit prices in the United States or even to treat pharmaceutical firms as regulated utilities. Return to Figure 17.4. It is theoretically possible to reduce price to P_R, raising quantity to Q_R, and enabling the firm to just cover its costs and earn a normal rate of return. However, the administrative complexity of regulating prices of multiproduct firms that are continuously introducing new products is enormous. Consider just the variations of a single product available in different strengths, forms (tablet, liquid), and delivery systems (oral, intravenous, patch, inhaler). In addition, Abbott (1995) found that pharmaceutical firms often set much higher introductory prices under regulation.

The complexity and potentially perverse effects of regulation prompt many analysts to conclude that cost containment is better left to private initiatives. They also worry about the damaging effects of price or rate-of-return controls on innovation. We cover both topics later in this chapter.

Competition and Generic Entry

Once a patent expires, other firms can enter the market. Entry barriers are considerably lower than those for new product development, and the 1984 Drug Price Competition and Patent Restoration Act further eased requirements. The act replaced the safety and efficacy testing under the 1962 amendments with much less costly bioequivalence tests. If the generic is approved, the FDA certifies it as "therapeutically equivalent" to the branded version. The new legislation has greatly increased generic applications, and as one would expect, firms target those markets with the greatest opportunities, in particular large markets and those where drugs treat chronic conditions (Bae, 1997).

What happens to prices and market shares after generic entry? Wiggins and Maness (2004) estimated an 83 percent drop in prices of anti-infectives (e.g., penicillins, tetracyclines) as the number of sellers increases from 1 to between 6 and 15, with further drops in price as more firms enter the market. This conventional finding on the impact of entry runs counter to a more complex story that had been developed for pharmaceutical pricing. Previously, Grabowski and Vernon (1992) examined 18 drugs that first experienced generic competition after the 1984 act. Generics captured one-half of their markets within two years. Surprisingly, though, as generic prices were falling, brand producers were raising theirs and widening the price gap over time. This phenomenon, corroborated by Frank and Salkever (1997), suggested that generics were not viewed as close substitutes by some patients or their providers. Pioneer firms can retain some monopoly power by capitalizing on the brand loyalty and relatively inelastic demand of this group. That is, as generics siphon off price-sensitive patients, the price-insensitive ones are left. The pioneer takes advantage of this market segmentation by raising brand-name prices for its loyal customers. (Pioneers can even introduce their own generic versions to compete in the generic segment.) We will return to the demand for generic substitutes in the section on cost containment.

Research and Development (R&D) and Innovation

Estimates of the drug industry's spending on R&D vary widely (Golec and Vernon, 2007), but there is no doubting the large amounts. Domestic R&D expenditures for members of the Pharmaceutical Research & Manufacturers of America (research-based pharmaceutical firms) rose from just over $1.5 billion in 1980 to $41.1 billion in 2015 (PhRMA, 2015), with another $10.1 billion spent abroad. Between 1980 and 1988, their share of domestic sales devoted to domestic R&D increased from 13.1 to 18.3 percent, stabilizing since then in the

range of 18–23 percent. Firms would not undertake these investments unless they could gain protection for their work. The patent system provides one method for providing protection.

A patent gives the holder the right "to exclude others from making, using, or selling the invention." The usual term for a patent is 20 years, but there are exceptions for drug products through exclusive rights granted by the FDA. To offset partially the delays posed by the long testing and regulatory review period, the Drug Price Competition and Patent Restoration Act of 1984 allows extensions for up to five years so long as the total effective patent life does not exceed 14 years. In 1983, Congress also passed the Orphan Drug Act, permitting extensions (and providing other benefits) for drugs designed to treat rare conditions that might otherwise not be profitable.[5] The exclusive right granted to KV Pharmaceutical, described in Box 17.1, falls under this Act.

On the one hand, patents and other legal protections, such as trademarks and copyrights, can lead to monopoly power, which is not in the public interest. On the other hand, a firm would be much less willing to expend millions of dollars on research if others can become free riders by mimicking its innovations. By being first, firms still would invest in R&D but at reduced levels. The question of just how much less is clearly an empirical issue.

Mansfield (1986) estimated that 60 percent of pharmaceutical drugs between 1981 and 1983 would not have been developed without patent protection. This figure is especially dramatic compared to the 11 other industries he sampled. The chemical industry was the only other industry with a substantial impact (30 percent).

Although patent protection has relatively small impacts on most industries, it is critical to pharmaceutical innovation. We, therefore, turn our attention to the determinants of pharmaceutical R&D and a conceptual framework to examine the effects of FDA regulations and patent law on innovations.

Investment Decisions

Net present value analysis provides a simple yet powerful approach to investment decisions. Letting R_t and C_t *r*epresent the revenues and costs in time, t, the net present value NPV of a project is given by:

$$NPV = \sum_{t=1}^{t=T} (R_t - C_t) / (1 + r)^t \tag{17.3}$$

where r is the discount rate or cost of capital and T is the life of the project. Following some of the discussion in the appendix to Chapter 4, under the standard decision rule, a project is accepted if the net present value is positive.

Several characteristics of pharmaceutical R&D become apparent if we break NPV into three components representing:

a The research, testing, and review period (m years)
b The effective period of patent protection (n years) after the product is launched
c The period following patent expiration (s years, where $m + n + s = T$)

$$NPV = \underbrace{\sum_{t=1}^{t=m} (R_t - C_t) / (1 + r)^t}_{\textbf{a.}} + \underbrace{\sum_{t=m+1}^{t=m+n} (R_t - C_t)/(1 + r)^t}_{\textbf{b.}} + \underbrace{\sum_{t=m+n+1}^{t=m+n+s} (R_t - C_t)/(1 + r)^t}_{\textbf{c.}} \tag{17.4}$$

In the first component, **a**, the firm will not have any revenue and there will be large, negative net cash flows reflecting the high R&D costs.

To offset these costs, a successful project will require even larger positive net cash flows in later years, particularly over the period represented by the second component, **b**, the effective patent life. The ability to charge high prices and/or reach large potential markets will have strong positive effects on *NPV*. Despite potential competition from generics and significant erosion of sales, products still may capitalize on brand recognition, marketing efforts, and new uses to remain successful after patent expiration, the last component, **c**.

This framework further tells us that regulations and testing procedures that increase costs in the first component reduce *NPV* and make an investment less attractive. Similarly, *NPV* is reduced by an increase in the length of the research, testing, and review period, because it must reduce the length of the patent protection period **b**. Conversely, changes such as reduced regulation or fast-track laws to lower initial costs and speed up the review process as well as extensions of patent rights each serve to increase *NPV*.

Finally, the risks are important. Projects with higher risks should be discounted at a higher rate or, put another way, high-risk projects need a high rate of return to be viable. To the extent that a firm can reduce risks, for example, by supporting a portfolio of diverse projects or sharing risks through joint ventures with other firms, the discount rate, *r*, diminishes and the likelihood of investment is increased. This also suggests that large firms have an advantage over smaller firms in R&D.

R&D Spending

We have noted the substantial total industry spending on R&D (as well as the slowdown in FDA approvals in recent years). Firm-level analyses of R&D provide some startling figures on costs and their recent growth rates. Focusing on the more significant innovations, DiMasi and colleagues (1991) estimated total costs, computed as capitalized expected costs and discounted at 9 percent, at $231 million in 1987 dollars per new chemical entity that was marketed. Because there is substantial attrition as projects move to successive stages of development, about two-thirds of the cost is attributable to the preclinical phase. In a controversial update covering the late 1990s, DiMasi, Hansen, and Grabowski (2003) estimated average out-of-pocket R&D costs for new chemical entities at $403 million, in year 2000 dollars. This figure reaches $802 million when capitalized at 11 percent.[6] Although an accompanying editorial by Frank (2003) supported the study's high quality, DiMasi's findings were attacked even prior to their formal publication. Two former editors of the prestigious *New England Journal of Medicine* (Relman and Angell, 2002) raised serious questions about the innovativeness of the pharmaceutical industry and many of its marketing practices. They further argued that DiMasi misrepresents R&D costs for new drugs, in that new chemical entities account for a minority of newly approved drugs. More recently, Light and Warburton (2005, 2011) have voiced concern about biases and other limitations of the proprietary and confidential survey data used by DiMasi. The accuracy and consistency of such data cannot be independently verified—an important caveat if one believes that pharmaceutical firms have an incentive to overstate development costs.

How do firms recover these formidable amounts? Our understanding of the returns to investment has been aided greatly by the work of Grabowski and Vernon (1994, 1996), which follows sales over the life cycle of a product. Grabowski and Vernon found that a product has an effective patent life of about 9 to 13 years and a market life of about 20 years. Cash flows do not become positive until the third year after launch, and sales peak in the tenth or eleventh year. The most significant finding is that a substantial portion of a company's revenue and profits comes from a few big winners. Only the top 20 percent of new drugs have substantially positive *NPV*; the *NPV* of the representative new drug is actually negative.

The highly skewed distribution of returns to investment reinforces the firm's need to diversify by having a large number of drugs under development to reduce risks. It also suggests difficulties for smaller firms that cannot sustain large R&D programs.

Firm Size and Innovation

Henderson and Cockburn (1996) ask: "Are the research efforts of larger firms more productive than those of smaller rivals, and if so, why?" They search for evidence of the effects of size on "important" patents granted and find that the returns to size are significant. This result is not surprising, but their contribution lies in distinguishing between "economies of scale" and "economies of scope." Recall that a firm experiences economies of scale when its long-run average costs decrease with higher output. It has economies of scope if the cost of producing two or more different products is less than the costs of separate production. Earlier work, which did not distinguish between these effects, showed mixed results on the effects of firm size on innovation.

For innovation, these concepts need clarification in that output is not easily or uniquely measured. One approach measures output through patents. Economies of scale arise if patents granted increase more rapidly than the size of an R&D program. Economies of scope occur if a joint program to develop different drugs is more productive than distinct programs.

Henderson and Cockburn found that size is important and that larger programs are more productive. However, it is due more to economies of scope than to economies of scale. There is little gain in increasing the size of an individual program beyond some minimal threshold level. Economies of scope arise when different research activities can share inputs, thus lowering cost. More important, spillovers of knowledge among projects raise overall productivity. The authors also found significant spillovers of information among firms.

More recently, Comanor and Scherer (2013) try to make sense of the wave of large mergers since the late 1990s that has coincided with a period of declining innovation (see also Box 17.1). Contrary to beliefs that mergers between large firms reverse the declining productivity, the authors argue that such mergers, by pruning "centers of initiative and decision-making" probably decrease the "chance that new technological prospects will gain large-scale support" (p. 113).

In addition to mergers, industry also is responding in other ways to raise the productivity of its R&D spending. Strategic alliances have risen sharply. The alliances, often between pharmaceutical and biotech firms, seek to pool efforts to innovate or bring products to the market more successfully. As part of this process, less pharmaceutical R&D is being conducted "in house" as firms look for R&D partners to reduce costs and spread risks. These partners include contract research organizations that are used to conduct clinical trials and, in some cases, to provide a broad range of drug-development services.

Prices, Price Regulation, and Innovation

Arguably, no issue is more important to drug policy than the effects of prices on innovation and, by implication, the effects of drug price regulation on innovation and the availability of drugs. Price regulation is often proposed as a means of limiting expenditures on drugs and, as we will describe, many other countries have adopted price controls or other forms of regulation. Should the United States rely largely on markets, even if imperfectly competitive, to determine drug prices and R&D activity, or is there an important policy role for price controls?

The theoretical framework represented by equation (17.4) suggests that higher drug prices and larger potential markets should spur R&D and consequently the rate of innovation.

Research provides strong evidence supporting these predictions. Vernon (2005) estimates that a price control policy that would lower pre-tax pharmaceutical profit margins to the average of those in non-U.S. markets would lower industry R&D investment by between 23 and 33 percent. Other studies find a consistent and substantial direct relationship between higher real drug prices and increased innovation.[7]

In analyses of drug launches, Danzon and colleagues (2005) investigated the number of launches and launch delays for 85 important new drugs in 25 countries over the period 1994–1998. The United States led all countries with 73 launches (Japan was lowest with 13). Higher expected prices and greater market size increase the number of launches and reduce launch delays.

The literature is clear on the adverse effects of price regulation on R&D investment, innovation, access to new drugs, and delays in availability. Are there benefits from regulation, such as increased access due to lower prices and reductions in expenditures, which could offset these adverse effects? Or, better yet, are there structural changes to the drug industry that could promote marginal cost pricing while maintaining high rates of innovation? There is considerable ongoing effort to answer these important questions.

Cost Containment

The rapid growth in drug expenditures has led to great policy interest in cost containment. President Clinton's proposed 1993 health care reform plan included a mechanism to regulate prices through caps geared to prices in other countries and to producer costs. Despite questionable methods, a series of government reports in the early 1990s, indicating that U.S. prices were higher than prices in Canada and the United Kingdom, intensified interest in drug price controls.

We already have addressed some of the difficulties of regulating prices for large numbers of constantly changing products. We also have addressed the problems of recovering common costs. If prices are driven toward marginal costs, R&D investment and the resulting innovation and access to new drugs may decrease, to the detriment of the public's welfare. Many countries have introduced various forms of regulation to rein in spending. Western European countries tend to control either producers' prices or reimbursement rates while England has a profit control system. In the United States, Medicaid programs "discount" prices and impose restrictions on utilization. The federal government also discounts prices for drugs purchased by the Department of Veterans Affairs and other federal agencies.

U.S. consumers purchase most outpatient drugs either out-of-pocket or with private insurance (Medicare's coverage for outpatient prescriptions started only in 2006). We, therefore, turn to the cost-containment efforts of the private sector, especially those introduced through managed care. These efforts include price discounting and the exercise of monopsony power, much like their public insurance counterparts. To narrow our discussion, we will describe three other strategies: higher copayments (often through multiple tiers of cost sharing), use of generic drugs, and the adoption of drug formularies.

Copayments

A higher copayment seems simple and straightforward, intended to shift a larger share of the cost burden to the patient and to decrease consumption of marginally beneficial drugs.

Moreover, under many current copayment plans, prescription drugs cost only a small fraction of their over-the-counter substitutes.

A higher copayment may also produce other results. Recall from Figure 17.2 that an increase in the out-of-pocket cost of a drug will decrease its use as patients substitute other inputs for drugs.

With no substitutability, use of drugs remains unchanged; only the distribution of the costs between patients and insurers is affected. As we noted earlier in the chapter, if significant substitutability among treatments is possible, the total health care cost could increase as inefficient cost-minimization combinations are selected.

However, when generic (multiple-source drugs) substitutes are available, strategies involving differential copayments for brand products offer the potential for substantial shifts toward lower-priced generics without affecting overall utilization. Many managed care plans now have three or more tiers of cost sharing in which the patient pays, for example, $7 for a generic drug, $15 for a brand-name product on an approved list (formulary), and more if it is not found on the formulary.[8] Patients and providers who consider the generic a close substitute will choose the generic.

The key question is whether there will be a large substitution toward generics. Motheral and Henderson (1999) examined two plans with tiered systems that increased brand-name copayments more than copayments for generics. They found little effect on total drug utilization. However, utilization of brand-name products decreased about 18 percent relative to a control group that had no price increases. This substitution produced substantial savings to the health plans.

Other evidence for a wider set of drug benefit arrangements is even more revealing. From a sample of over 400,000 working-age adults, Joyce and colleagues (2002) found that doubling copayments for all drugs from $5 to $10 reduced average drug spending by 22 percent, and by about one-third when copayments were doubled in two-tier plans. Adding a second $20 tier for brand-name drugs that previously had a $10 copayment for all drug purchases reduced spending by 19 percent. Drug spending was also reduced by 8 percent in two-tier plans that mandated generic substitution as compared to those that did not mandate such substitution. Numerous analyses of the Medicare Part D plan (discussed in Chapter 20) show that 100 percent copayments in the so-called "doughnut hole" lead to substitution of generic drugs for brand-name drugs, and on occasion for the patient to reduce frequency or even stop taking the drugs.

Finally, we caution that higher copayments and other cost-sharing schemes not only reduce costs, but also change treatment. Goldman et al. (2004) found substantial decreases in utilization within the most common drug classes from a doubling of copayments. Reductions ranged from a low of 25 percent for antidiabetics to highs of 44 percent for antihistamines and 45 percent for nonsteroidal anti-inflammatory drugs (NSAIDs). Utilization for those with chronic illnesses was less responsive to the copay increases. Nevertheless, the authors were concerned about the health effects of the large reductions, especially for patients with diabetes.

The cost-sharing literature reinforces this concern. Gibson and colleagues (2005) concluded that these arrangements generally work as intended—by encouraging generic use and limiting overuse. But their study also found reports that higher cost sharing can also disrupt treatment through lower levels of adherence, lower use of essential medicines, and, in some cases, drug discontinuation.

Philipson and colleagues (2010) highlight the potential for such problems by studying patients with acute coronary syndrome who underwent stent implantation. Patients with high cost-sharing were less likely to take antiplatelet drugs (a highly effective therapy) following the implantation and more likely to discontinue their use within the first year. As a result, these patients experienced worse outcomes and had higher total costs due to increased re-hospitalization.

Generic Substitutes

With the expiration of patents on some important drugs and the cost-containment efforts made by many plans, about 83 percent of the prescriptions written in 2015 were filled with unbranded generic drugs—up from 50 percent in 2005. The percentage of generics for multiple-source products was likely to be much higher. Many top-selling drugs now have generics. With generic prices considerably below their brand-name counterparts and FDA certification of their therapeutic equivalence, one would expect little resistance to generic versions from physicians and their patients. This has not always been so.

Interest in promoting generic products has a long history. Most states passed antisubstitution laws after World War II, prohibiting pharmacists from substituting a generic for a prescribed brand, but mounting cost pressures led to reforms. In 1970, Massachusetts became the first state to legalize drug product selection and, by the end of the decade, most others followed. Despite the potential for cost-saving, an early study of Michigan's substitution laws was revealing (Goldberg et al., 1979). After 1974, substitution by a pharmacist was allowed unless the doctor wrote "dispense as written" or "DAW," but physicians wrote relatively few prescriptions with this restriction. However, pharmacists provided substitutions for less than 2 percent of all multiple-source prescriptions.

Substitution has increased well beyond the levels of the 1970s, but stood at just 19 percent in 1984, when the Hatch-Waxman Act of 1984 passed. The legislation allows a generic producer to file an abbreviated new drug application and to use the pioneer company's clinical research. Since then, efforts by managed care and other third-party payers have greatly increased generics' share of the prescription drug market.

Therapeutic drug substitution, a practice widely opposed by physician organizations, is a different though related phenomenon. Therapeutic substitution would replace the prescribed drug with a chemically different drug from the same drug class that is expected to produce equivalent clinical effects. Johansen and Richardson (2016) found that spending on branded drugs could be substantially reduced under therapeutic substitution with much of the savings concentrated in a small number of drug classes.

Drug Formularies

Managed care's strong financial interest in cost containment has led to policies that go well beyond copayment strategies to promote generics. Many plans monitor physicians and require substitution when generics are available. Many also use pharmacy benefit managers to negotiate discounts and improve the efficiency of their claims-processing and pharmacy operations. They are increasingly adopting other methods such as drug-utilization review programs and lower-cost, mail-order sources for prescriptions. One of the most ambitious and controversial strategies, however, involves the use of formulary committees to develop a list of approved drugs. A positive formulary restricts the choice of drugs to those on the list. A negative formulary excludes drugs on a list.

The formulary review and approval process can be elaborate, dealing not only with generic substitution but also with recommendations of different drugs to treat a condition. Ideally, the review committee will periodically review all drugs in each therapeutic class for their clinical effectiveness, safety, and cost. Mather (1999) observes that, when properly implemented, the formulary can be an effective and well-accepted tool. If drug-product decisions are based largely on cost instead of clinical outcomes, Mather suggests that "the health system may experience higher overall costs and the pharmacy benefit may be sharply criticized by health plan providers, enrollees, and suppliers wishing to see their products on the list" (p. 277).

The widespread adoption of formularies has elevated the importance of pharmacoeconomic analyses. Pharmaceutical firms face strong pressure to provide clinical and economic data that justify their inclusion in the formulary. Formulary committee members require the skills to compare, interpret, and analyze data from many different sources and to justify their decisions. They are under especially great pressure from both patients and providers when new and improved drugs come onto the market.

Reference Pricing

Reference pricing is not new but it has received considerable attention in recent years. There are two forms of reference pricing. *External reference pricing* involves international price comparisons that some countries use as a benchmark to control drug prices or negotiate prices for products protected by patents. Ruggeri and Nolte (2013) describe the variations in the strategies of those highly developed countries (e.g., Canada, France, and Germany) that have adopted external reference pricing.

Internal reference pricing is used within countries. Here, a third-party payer establishes the reimbursement rate, or reference price, for drugs within a therapeutic class. It will often be the lowest priced drug in that class (Lee et al., 2012). Because patients are responsible for the difference between the reference price and the price of the drug, firms have an incentive to limit prices.

In their review of the recent literature, Lee and colleagues concluded (p. e434) that reference pricing "reduced drug prices, increased utilization of and adherence to target drugs, and promoted switching behavior from expensive products to alternatives at or below the reference price." They also found that it significantly reduced patients' costs as well as overall payer costs. Their review suggests that reference pricing should be more widely considered for the United States which has only limited experience with this strategy.

New Drugs and Health Care Spending

It might appear from the material presented that drug spending in the United States has gone beyond any level that analysts might consider reasonable or efficient. If so, then the United States is not alone among developed countries. As measured by the percent of health spending devoted to drugs, the United States actually ranks below many other industrial countries (though it is the biggest spender in absolute dollars). Many factors account for the wide variations in shares across countries in addition to price differences. These include differences in physician practice patterns, cultural attitudes toward drugs, licensure laws, and insurance and other financial arrangements (physicians are permitted to dispense drugs in some countries). All nations are dealing with the same pressures and the need to find mechanisms to promote cost-efficient use of drugs and other health care inputs. After all, the most important economic question is the value of drugs in producing health.

Fortunately, economists have addressed this question. Frech and Miller (1999) took advantage of the wide differences in drug utilization and health status across 21 OECD countries to investigate the productivity of drugs, other medical inputs, wealth, and lifestyle. They used regression analysis with alternative measures of a nation's health status as the dependent variable. Independent variables included a country's wealth, pharmaceutical spending per capita, other health spending per capita, and various lifestyle indicators.

Some of the findings are consistent with our analyses in previous chapters. Wealth, measured by GDP per capita, showed significant and positive effects on life expectancy. A doubling

of GDP increases life expectancy by 6 percent at age 40 and by 9 percent at age 60. The consumption of animal fat, the most important lifestyle variable, has positive effects on life expectancy at birth, age 40, and age 60 up to certain levels of consumption (92, 78, and 70 grams per day, respectively). Nonpharmaceutical inputs showed no statistically significant effects at any of the three age levels.

Pharmaceutical spending per capita has strong positive effects on life expectancy. A doubling of drug spending increases life expectancy by 2 percent at age 40 and 4 percent at age 60. The lifetime pharmaceutical cost of extending life by one year in the United States for males and females at age 40 are $21,000 and $23,000, and about $17,000 and $19,000 at age 60.

Subsequent research generally showed that newer drugs are well worth the extra cost. In Lichtenberg's (2007a) analysis of individual medical conditions, reducing the age of drugs used in treatment reduces nondrug spending by a factor of 7.2 (8.3 for the Medicare population) relative to the increased drug spending. The bulk of the savings (i.e., offset effects) come from reduced hospitalization. Lichtenberg (2008) also examined data for 20 OECD countries specifically for cardiovascular disease. Cardiovascular disease is the leading cause of hospitalization and death in many nations. The adoption of newer drugs reduced the cardiovascular age-adjusted mortality rate and number of hospital days (through lower hospitalization rates and lengths of stay). The per capita savings on hospital stays ($89 in 2004) from adoption of the new drugs was 3.7 times as large as the increase in per capita spending for cardiovascular drugs ($24).[9]

The ACA and the Pharmaceutical Industry

The Affordable Care Act (ACA) has nearly 50 provisions that directly impact the pharmaceutical industry (Milne and Kaitin, 2010). Some are discussed in Chapter 22 while others are too narrowly focused to be covered here. At this point, we note that the pharmaceuticals face two significant and costly provisions. First, an annual fee is imposed on makers of brand-name drugs. The fee is determined through a complex formula but it began at $2.5 billion in 2011. Second, seniors who are enrolled in a Medicare prescription drug plan receive substantial discounts on brand-name drugs and lesser discounts on generics when they reach the "doughnut hole." (The "doughnut hole" is the range over which drug coverage ends and before catastrophic coverage begins.) These discounts cost the industry about $2 billion per year.

What about the gains? One of the biggest benefits is not immediately obvious. Many supporters of health care reform in the United States favored a bill that would have allowed the Department of Health and Human Services to negotiate the prices of all drugs purchased by Medicare beneficiaries. In exchange for the industry's support of reform, and to gain support from a major organization representing the elderly, Congress dropped any such provision from the legislation that President Obama ultimately signed in 2010. In other words, the industry dodged a major threat to its profitability.

Other benefits include grants for projects that can lead to major advances in basic research and to "high need cures," as well as tax credits to smaller companies to develop new therapies. The largest and most obvious gain, however, comes from the millions of previously uninsured consumers that will be covered by private insurance or an expanded Medicaid program. According to one estimate, the ACA will contribute $115 billion in sales and $45 billion in additional profits over the coming decade.[10] The industry appears to be a net winner under the ACA.

Conclusions

Our overview of the pharmaceutical industry focused on individual provider and patient decisions, firm pricing and investment decisions, and public policy issues in regulation and cost containment. Throughout, we have emphasized the use of basic economic tools to simplify and make sense of complex problems.

Pharmaceutical companies and their profits are highly visible and they make convenient targets. However, newer drugs appear to lower mortality and produce net reductions in health care spending. Careful research thus cautions against ill-conceived proposals to limit the availability of drugs and to contain drug spending.

The drug industry is experiencing profound change as a result of mergers between large firms, the declining productivity of its R&D, and the effects of many ACA provisions. The industry appears to be a "winner" under the ACA but it is still difficult to project whether "Big Pharma" will continue to thrive. Spending on prescription drugs has levelled off in recent years but the industry still has many critics and a negative public image. With cost control a never-ending challenge under health care reform, the industry will likely face continuing legislative attempt to control prices and limit its profitability.

Summary

1 In 2014, prescription drug spending accounted for 9.8 percent of national health care spending, up from 8.8 percent in 2000 and 4.7 percent in 1980.
2 The pharmaceutical industry is characterized by significant barriers to entry and substantial regulation. The industry has been regulated since 1906, but 1962 legislation had the most profound effects.
3 Pharmaceuticals substitute for health and nonhealth inputs in the production of health. A rational patient will select least-cost input combinations. Relative changes in a patient's out-of-pocket costs resulting from insurance will lead to substitution of drugs for other health care inputs or vice versa.
4 Technological change, even when the benefits are marginal, is often cost-increasing due to insurance.
5 Pharmaceutical firms earn higher-than-normal profits, but the extent of their profits is exaggerated by conventional accounting data.
6 Opportunities are substantial for price discrimination. Markets, or groups of buyers, with inelastic demand will pay higher prices.
7 Generic products often capture a significant share of the market following patent expiration. Nevertheless, through effective marketing and promotion, trade names still can retain a monopoly premium.
8 R&D spending is substantial, and a firm's profits often depend on a few big winners. Firms need to earn substantial amounts over the effective life of a patent to justify the risks.
9 Larger R&D programs are more productive than smaller programs. Economies of scope play an important role.
10 Price regulation reduces investment in R&D, the rate of innovation, and the number of drug launches.
11 Price regulation, rate-of-return regulation, national formularies, and reference pricing are used in other countries to contain costs. In the United States, higher copayments, the promotion of generic substitutes, and other managed care strategies, such as drug formularies, have been adopted.

12 Higher copayments encourage substitution toward generics as well as a decrease in overall utilization. In some cases, they may also lead to disruptions in treatment.
13 Spending on drugs in the United States is not out of line compared to other developed countries.
14 Drugs are highly productive in improving health compared to other medical inputs. New drugs reduce health care costs, largely through reduced hospitalization, relative to the additional cost of these drugs.
15 The pharmaceutical industry is affected by many provisions of the ACA. The revenues coming from the additional insured consumers are likely to more than offset the costs of the concessions made by the industry.

Discussion Questions

1 Explain and distinguish between the "concentration ratio" and the "HHI." What are the limitations of these measures within the context of the pharmaceutical industry?
2 Direct-to-consumer (DTC) advertising of prescription drugs has grown rapidly. List several products with which you have become familiar as a result of such advertising. Discuss the pros and cons of DTC advertising from the perspective of physicians and patients.
3 Use Figure 17.1 to explain how an isoquant can be positively sloped. Under what circumstances may a patient actually end up in the positively sloped region (e.g., at point *H*)?
4 Regulation is often proposed (and widely used in other countries) to limit prices or profits. Discuss possible adverse effects of regulating prescription prices. In light of your discussion, what accounts for the strong pressure in many countries, including the United States, to regulate prices?
5 What are barriers to entry? Describe three potential barriers in the pharmaceutical industry. What are some consequences of these barriers?
6 In 2004, Congressman Dennis Kucinich proposed the Free Market Drug Act. This legislation would have removed patent protection on drugs developed with public funds and given control over pharmaceutical R&D to the National Institutes of Health (NIH). Evaluate this type of proposal in terms of the effects on price, competition, and level of innovation.
7 There are wide differences across countries in the share of health resources spent on drugs. Describe possible economic and noneconomic factors that may contribute to the variation.
8 Direct-to-consumer advertising has been criticized for possibly misleading patients and for increasing spending on drugs. Discuss the benefits and costs of DTC advertising.
9 Media reports often show much higher drug prices in the United States than in other countries. Analyses by economists often show that the price differential is not as large. Describe some possible problems in comparing domestic with foreign drug prices.
10 Consider the information in Box 17.2. Are the CEOs of Turing, Valeant, or other firms seeking profitable opportunities by purchasing generics and subsequently raising prices just engaging in good business strategies? Are they improving social welfare in the sense that economists use this concept?

Exercises

1 Use Figure 17.2 to explain why cost minimization through a tangency between an isoquant and a budget line does not apply in cases where *D* and *M* are either perfect complements

or perfect substitutes. Describe the cost-minimization process and the effects of changes in the relative prices of D and M in these special cases.

2 Assume that a patient has 80 percent coverage for medical services but no coverage for prescription drugs. An 80 percent drug benefit is added. Show graphically what will happen to the relative utilization of M and D, and total spending on health care, to attain a given health status. What happens to the amount spent on drugs (insurance plus patient amounts)? Why will total spending on health care diminish when the 80 percent drug benefit is added?

3 A pharmaceutical firm faces the following monthly demands in the U.S. and Mexican markets for one of its patented drugs:

$$Q_{US} = 300{,}000 - 5{,}000\,P_{US} \text{ and } Q_X = 240{,}000 - 8{,}000\,P_X$$

where quantities represent the number of prescriptions. Assume that resale or arbitrage among markets is impossible and that marginal cost is constant at $2 per prescription in both markets. Monthly fixed costs are $1 million in the United States and $500,000 in Mexico.

(a) Draw the demand, marginal revenue, and marginal cost curves for each market. Estimate the profit-maximizing prices and quantities graphically and/or determine the solutions algebraically. What are the firm's total profits?

(b) Determine the quantity in each market and maximum possible total profits if the firm engages in perfect (first degree) price discrimination. Is this behavior possible?

4 Assume that the firm in Exercise 3 cannot prevent resale and is forced to set the same price in both markets. Find the price graphically and/or algebraically and show that total profits are less than those from part 3a.

5 For your answer in 3a:

(a) Calculate the price elasticity of demand in each market at the optimal price.

(b) Verify that the prices and elasticities are consistent with the profit-maximizing formula given in Footnote 4.

(c) Why are both elasticities fairly close to unity? (Hint: Think about the requirement for profit maximization when marginal cost is zero.)

(d) If a firm finds that its price elasticity is numerically less than 1, what advice would you have?

6 Consider only the U.S. market from Exercise 3. Graph solutions to parts 6a and 6b using the demand, average cost, and marginal cost curves. Also try to develop the answers algebraically.

(a) Price regulation is proposed. Find the regulated price that enables the firm to cover all its costs. Caution! There are two mathematical solutions. Which one will regulators prefer?

(b) Find also the economically efficient price (i.e., one that is consistent with marginal cost pricing). What subsidy per prescription is required to enable the firm to cover all its costs?

7 Compare your results in Exercises 6a and 6b with the profit-maximizing solution for the United States obtained in Exercise 3a. Explain which of the three alternatives you would prefer if you were responsible for public policy. Be sure to consider some of the problems of regulating prices.

8 Many insurance companies increased premiums by 15 percent or even more in the early years after 2000, blaming soaring pharmaceutical costs for their premium increases. Evaluate the validity of this justification. (Hint: Consider spending on pharmaceuticals as a share of national health expenditures.)

Notes

1 Thalidomide still is marketed with a warning on pregnancy. It is used to treat a skin condition caused by leprosy and to treat multiple myeloma, a cancer of plasma cells.
2 The legislation also gave the FDA authority over the manufacturing process and extended testing requirements to include generic drugs and drugs that are similar to available products ("me-too" drugs). It also transferred regulation of drug advertising for prescription products from the Federal Trade Commission to the FDA.
3 The terms *efficacy* or *effective*, as used by the FDA, mean that a drug has positive effects compared to a placebo. Efficacy does not imply that the product is cost-effective or that it meets other economic criteria for efficiency.
4 A convenient formula for profit maximization is:

$$P = \frac{MC}{\left[1 + \frac{1}{E_p}\right]}$$

where E_p is the algebraic value of the price elasticity of demand. With the same MC in both markets, price must be higher when demand is less elastic (e.g., –2 versus –4).
5 Schweitzer (1996) describes the incentives provided by the act and how they have led to a large number of products given orphan status. He argues, however, that the act does little to help populations in developing countries who suffer from diseases that are rare in the developed world. Together with the high cost of many existing drugs, the responsibility of industrialized nations and the pharmaceutical industry to poorer nations is an ongoing issue.
6 See also DiMasi and colleagues (2005) and the Tufts Center for the Study of Drug Development (http://csdd.tufts.edu/news/complete_story/internal_news, accessed April 21, 2011) for responses to these concerns. Citing updated work by Dimasi and colleagues (2014), the industry currently uses a value of $2.6 billion as the cost of developing a new drug (PhRMA, 2015) including "failures and capital costs."
7 The elasticity of innovation with respect to drug prices is about 0.6 (Giaccotto, Santerre, and Vernon, 2005, and Lichtenberg, 2007b).
8 Many Medicare Part D plans, and some commercial plans, have introduced four-tiered systems in which the fourth tier is reserved for certain expensive drugs including biologic drugs used to treat cancer. The fourth tier is handled as coinsurance; for example, the patient pays 25 percent of the prescription's cost with no limits, in some plans, on the out-of-pocket amounts. See also Box 8.2.
9 Civan and Köksal (2010) similarly found that newer drugs reduce total health care spending with the largest reduction occurring for hospital care. However, Law and Grépin (2010) describe some serious biases in the methodology used by Lichtenberg and others. The controversy over the offset effects of new drugs is not a completely settled issue.
10 Estimates by GlobalData as reported in "ObamaCare Will Bring Drug Industry $35 Billion in Profits," *Forbes*, May 25, 2013: forbes.com/sites/brucejapsen/2013/05/25/obamacare: Accessed February 27, 2014.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 18

Equity, Efficiency, and Need

In this chapter

- Efficiency and Competitive Markets
- Deviations from the Competitive Model in the Health Care Sector
- Promoting Competition in the Health Care Sector
- An Economic Efficiency Rationale for Social Health Insurance
- Need and Need-Based Distributions
- Horizontal Equity and Need
- Theories of Social Justice
- Conclusions

Are health care markets healthy? Are they efficient, and do they provide the care that people need? These questions occupy the minds of the many people who study health care. Efficiency questions arise because of the high costs that people must pay for health care. Are these costs too high? Likewise, equity questions occur because many people, certainly including the uninsured, face barriers in obtaining health care.

Compared to Canadians and Europeans, Americans will more likely find experts who favor competitive market solutions to health care system problems, though most here (as elsewhere) will argue that this approach is often ill-suited to the nature of health care markets. Canadians and Europeans are more willing to use government interventions, although the success of the Affordable Care Act (ACA) in 2010 suggests that the United States may have moved in that direction, albeit with some substantial political opposition. Similarly, Medicare reform remains centrally important, and many Americans find equity and efficiency in the health care system to be the fundamental issues.

A solid background in these issues requires a study of the economics of efficiency, the departures of many health care markets from the competitive model, the role of equity concerns, and issues of social justice theory. Because of the central role of "need" in health equity discussions, we must also investigate the meaning of health care need. These subjects are the themes of the present chapter.

The chapter focuses on welfare economics, the study of normative issues that bear on economics. "Normative issues" deal with how people believe the economic world *should be*, as opposed to "positive issues" that deal with how the world of economics functions *in practice*.

So, welfare economics would encompass those that are critical of existing markets and question the distribution of goods and services. Some health economists (Hurley, 2000; Culyer, 1989), however, dispute this understanding, arguing that an "extra-welfarist" viewpoint is required, rejecting some or all of the philosophical principles on which welfare economics is based. Yet other theorists find the tools within welfare economics to understand the concepts of welfare and efficiency that concern us the most (Absolo and Tsuchiya, 2004).

While we focus on standard welfare economic theory, we will explain sources so students can explore the extra-welfarist view more fully. We first describe the standard results for competitive markets, but also the many market flaws that cause markets to deviate from competition, causing many competitive efficiency propositions to fail. We will also explain and describe the role of need and need-based distributions in the health economy. Finally, we will present theories of social justice and explain why welfare economic claims must be grounded in a philosophical position on justice.[1]

Efficiency and Competitive Markets

We clarify the meaning of economic efficiency within the context of the Edgeworth box for exchange. This approach derives theorems in a graphical framework that theorists have also developed in more sophisticated mathematical models. The analysis here generates the First Fundamental Theorem of Welfare Economics, and illustrates the Second Fundamental Theorem as well. The First Theorem demonstrates that competitive markets under certain conditions are economically efficient. The Second Theorem establishes that a society can achieve any desired economically efficient outcome by competitive markets if it starts from the appropriate initial endowments.

The Concept of Pareto Efficiency (Optimality)

In the early twentieth century (1906), economist Vilfredo Pareto and his followers defined the concept of efficiency most frequently used by economists today. According to them, an economically efficient (optimal) outcome in society is one under which it is impossible to improve the lot of any person without hurting someone else. Pareto efficiency also implies that no further exchanges would be found that could improve the lot of everyone to some degree. An efficient economy necessarily would have exhausted all means for mutual gains.

The Edgeworth box, using a hypothetical two-person economy and showing exchanges between these two people, provides a context in which to make the idea of Pareto efficiency clear. The box also is convenient for describing the mutual gains from trade and for defining the Pareto concept of efficiency.

Suppose that persons A and B, say Abner and Belinda, inhabit a desert island, forming a two-person economy. Further suppose that only two goods are available on the island. Food, F, is gathered and is available in a fixed total amount, F_0, and medicine, M, is likewise available in a fixed amount, M_0.

To form the Edgeworth box, consider Figure 18.1. Abner's preference map (indifference curves) starts from the southwest corner. There is no reason to draw the axes out further than M_0 and F_0, which represent the total amounts of medicine and food available on the island. Belinda's preference map is similar to Abner's except that it starts at the northeast corner. It is also constrained by amounts M_0 and F_0.

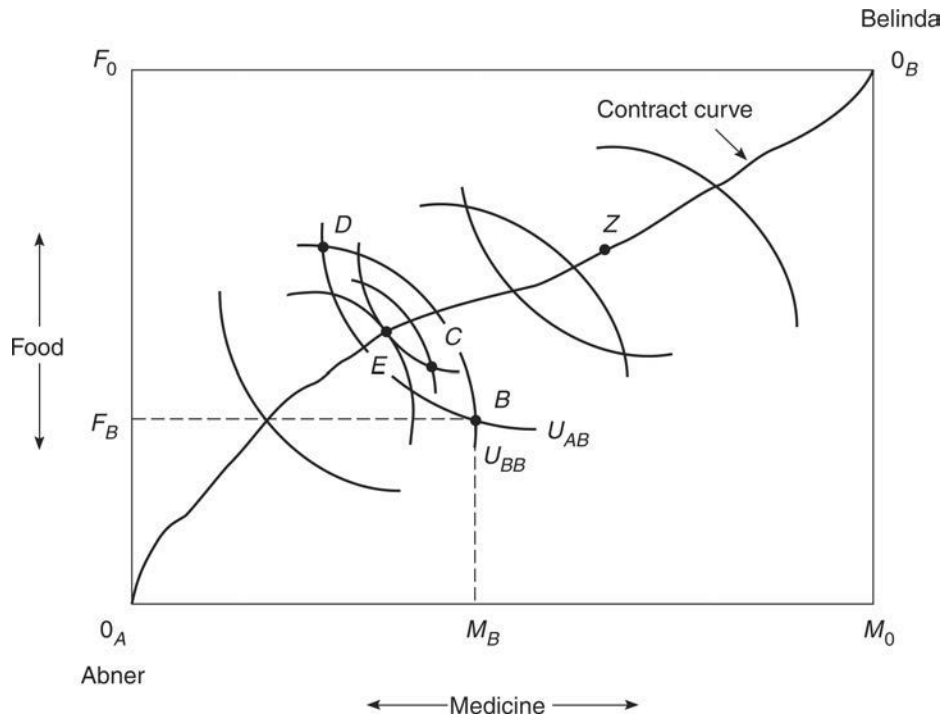


Figure 18.1 Edgeworth Box for Exchange; Pareto Efficient Points Lie on the Contract Curve

Any point in the box represents a complete and exhaustive distribution of the island's endowment of food and medicine. For example, point *B* represents a distribution in which Abner has M_B units of medicine and Belinda has $M_0 - M_B$ units of medicine. Similarly, at *B*, Abner has F_B units of food and Belinda has $F_0 - F_B$ units of food. With this orientation, we ask whether point *B* is an economically efficient distribution.

The answer to this question must be no. To see this, examine by comparison point *C*. Point *C* lies on an indifference curve that is above (to the northeast of) indifference curve U_{AB} and, therefore, *C* is superior to *B* in Abner's view. Similarly, point *C* lies on an indifference curve that is above (to the southwest of) indifference curve U_{BB} and, therefore, *C* is superior to *B* from Belinda's view. Because point *C* is attainable and improves the lot of both persons while harming neither, it follows that the original point *B* is not economically efficient.

Geometrically, we can repeat the analysis regarding point *B* for any point that forms a "lens" from the indifference curves passing through it. A lens is formed by the indifference curves U_{AB} and U_{BB} from point *B* to point *D*. Whenever we can find such a lens, we can identify one or more other points superior to the initial point. Reapplying this reasoning, point *C* is also not Pareto efficient. Pareto-superior moves, where the welfare of both improves, can also be made from point *C*. In contrast, a Pareto-efficient point in the box is a point of tangency between two indifference curves, such as point *E*. It is impossible to move from a point of tangency without harming the lot of one of the two persons.

Each of Abner's indifference curves will have a point of tangency with one of Belinda's indifference curves. We call the collection of all Pareto-efficient points in the box the *contract curve*, which is so labeled in the figure. For example, at point 0_A, Belinda has all of both goods, and even if many or most people consider this inequitable, it *is* Pareto efficient because giving any of either good to Abner would make Belinda worse off.

Trading along the Budget Line

Having defined efficiency in the context of the Edgeworth box, we next ask whether the competitive market generates an efficient equilibrium in exchange. In a competitive market, each person treats prices as given and responds to prices by choosing the utility-maximizing bundle subject to his or her resource constraint. The resource constraint depends on the person's initial endowment of food and medical care. Let point *V* in Figure 18.2 represent the initial endowment for this two-person economy. Either person may trade away from his or her initial endowment at the market prices. Thus, Abner's resource constraint will be represented by a budget line passing through point *V*. As with any budget line, the slope of this line is the negative of the ratio of the price of medical care to food.

The slope of the budget line represents the rate at which one can trade one good for another at market prices. The steeper the budget line is in Figure 18.2, the greater the price is of medical care relative to food. For example, budget line *AB* represents a relatively lower price of medicine relative to food than does budget line *CD*.

The Competitive Equilibrium

To find the competitive equilibrium, we must identify how much each person would be willing to trade. Abner's offer curve, for example, is the collection of points representing his offer for trade at each possible set of prices. Start at point *V*. Given budget line *AB*, Abner stays at point *V*, the point of tangency between budget line *AB* and the highest indifference curve that is attainable.

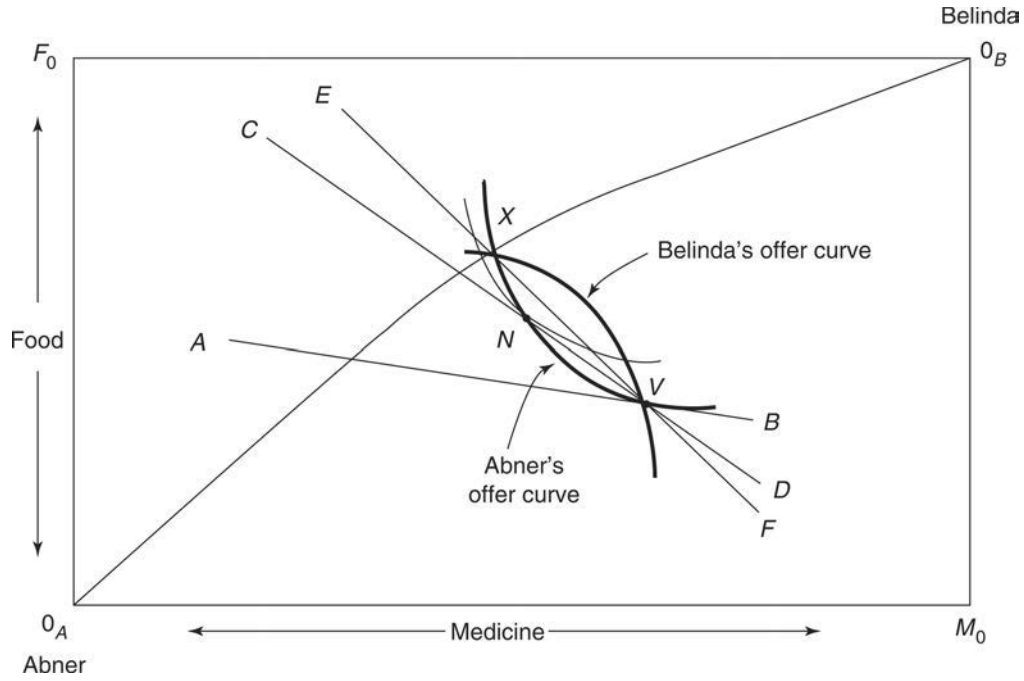


Figure 18.2 The Intersection of Offer Curves Determines the Competitive Equilibrium

Suppose the price of medical care were higher relative to food, leading to the steeper budget line *CD*. Given budget line *CD*, Abner would trade some medicine for some food to go from point *V* to point *N*. With budget line *EF*, Abner would trade to point *X*. Connecting all such points generates Abner's offer curve.

The figure also shows Belinda's offer curve, beginning at endowment point *V*. The two heavily shaded offer curves represent voluntary trades for the two parties. For trade, as in a competitive market, to be mutually voluntary, the offers of the two persons must agree. The offer curves agree only at their point of intersection, labeled point *X* in the figure. Point *X* thus constitutes the competitive market equilibrium in exchange for this two-person economy, starting with the endowment of *V*.

The First Fundamental Theorem of Welfare Economics

Is the competitive equilibrium, *X*, Pareto efficient? Yes, and there are two reasons that it must be so for every competitive equilibrium. The intersection of two offer curves represents a trade made at competitive prices starting at point *V*. Each person is at a point of tangency between the budget line and the highest attainable indifference curve. At point *X*, Abner's indifference curve (not shown) is tangent to the budget line. Likewise, at point *X*, Belinda's indifference curve (also not shown) is tangent to the budget line. Because these indifference curves are tangent to the same budget line at the same point, they must be tangent to each other. Because they are tangent to each other at point *X*, this point is Pareto efficient. The same argument applies for any competitive equilibrium; therefore, we have

shown the First Fundamental Theorem in this context, namely that the perfectly competitive market equilibrium is Pareto efficient.

The theorem makes the competitive market solution attractive. If we can establish perfect competition, then the market forces left to their own workings will generate an efficient outcome—an invisible-hand solution. However, the theorem evokes several serious questions: Can we achieve competitive markets in health care? Is the context of this theorem appropriate for health care? Would the competitive market solution be equitable or would it leave too many people without adequate health care? We will address each of these questions. However, we begin this process by exploring the issue of equity within the context of the Second Fundamental Theorem.

Redistribution of the Endowment

We extend the applicability of the First Fundamental Theorem with the Second Fundamental Theorem, which states that given an appropriate endowment, any Pareto efficient outcome can in principle be achieved by a competitive market. Figure 18.3 illustrates the significance of this theorem.

In Figure 18.3, suppose that the initial endowment is *V*, and suppose that this endowment results in the competitive outcome represented by point *E*. Point *E* is only one of an infinite number of Pareto efficient points. It may be an outcome that many view as inequitable, here either Abner or Belinda, or both. In real life, the society may have millions of members, and plausibly a majority of people may perceive this market outcome to be inequitable.

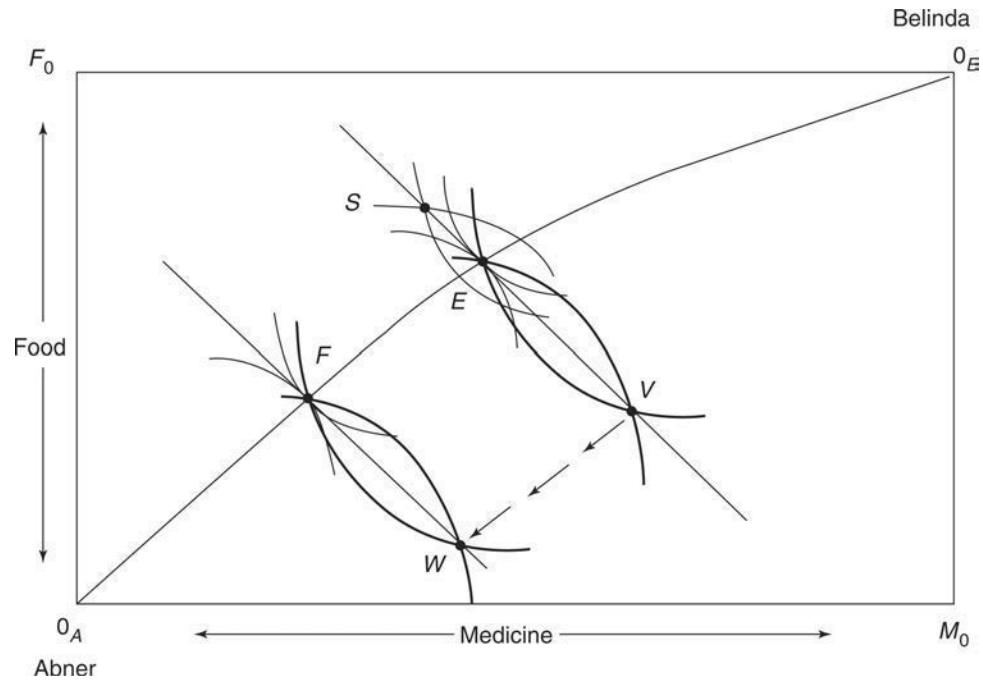


Figure 18.3 A Preferred Efficient Outcome May Be Achieved by an Initial Redistribution of Resources

The Second Theorem, however, defines a central role for competitive markets, even in cases where many view some of the competitive outcomes as inequitable. Suppose, in Figure 18.3, that society prefers outcomes in the vicinity of point *F* to outcome *E*. By the Second Theorem, a competitive market can achieve the desired outcome, but it requires a different initial endowment from point *V*. As shown, the endowment point *W* is a suitable point from which to achieve an equitable market outcome, point *F*. Seen this way, redistribution combined with competitive markets generates an efficient and equitable outcome. This contrasts with command systems that reject free markets, as well as with alternative schemes, such as price discrimination.

Price Discrimination

Some propose achieving a more equitable outcome by providing certain services to the poor at reduced, subsidized prices. Readers may be surprised to learn that such systems are not consistent with Pareto efficiency. Consider the proof of the efficiency of competitive markets. It was crucial that both parties achieve a point of tangency to the same budget line. If the poor are charged different prices than the rich, the two groups face different slopes of their budget lines. The result would be a position such as point *S* in Figure 18.3, a point that is not Pareto efficient.

For an intuitive argument, a program subsidizes the poor in purchasing bread. The poor will adapt to the subsidized price until the rate at which they were willing to trade bread for other goods equals the rate at which they could exchange the goods at the subsidized price. The result is that the poor will undervalue bread in comparison to the wealthy. It will be more efficient for the poor to buy up bread and sell it to the rich. Such a side market, which would improve efficiency in the bread example, is not possible for many forms of medical care (treatment for broken legs, for example), which are not easily transferrable. Thus, subsidized prices for medical care will likely generate an inefficient equilibrium.

The two theorems, along with the inefficiency of price discrimination, suggest the superiority of income transfers as a solution to equity problems in health care markets. In Figure 18.3, the following situation takes place: Transferring initial resources between the two persons and then allowing the market to work will achieve an efficient outcome within the equitable range.

Trade-Offs between Equity and Efficiency

The theoretical superiority of redistribution of income to programs, such as price subsidies, has led many analysts to favor income maintenance programs as policy tools to offset the problems of poverty, including the problems of access to health care. Income maintenance programs are government programs designed to provide cash subsidies to the poor to maintain their incomes at or above a preset floor. Despite continuing interest in such programs, policymakers often have hesitated to use large-scale income redistribution.

Economists explain a major criticism of income maintenance by appealing to Okun's (1975) analogy of the leaky bucket. The act of transferring wealth from one group to another in society may generate disincentives that discourage productive effort. The taxpaying group incurs a tax burden that may reduce work incentives, and the recipient group receives subsidies that may reduce incentives to work and to self-help. By analogy, when we transfer income, our task is similar to transferring water in a leaky bucket. The amount of income available for redistribution may decline as a result.

Blank (2002) challenges the equity–efficiency trade-off idea, arguing that situations exist where the efficiency costs of improving equity may be very small, such as when the group receiving the benefit is unlikely to change its behaviors. She further posits that in some cases equity and efficiency are complementary.

In the 1960s and 1970s, the federal government sponsored large-scale experiments to investigate the degree of work loss induced by the incentives inherent in income maintenance programs. These experiments reported reductions in work effort on average of between 5 and 10 percent. However, the work reduction estimates were considerably higher for certain subgroups, such as "male nonheads (of families)" and women. Also, results generated in an experimental situation make it difficult to predict the results if the program were to become universal and permanent.

Deviations from the Competitive Model in the Health Care Sector

Another major criticism of the applicability of our theoretical analysis concerns the question of whether health care markets are sufficiently competitive or whether we can make them sufficiently competitive to obtain competitive outcomes. Substantial differences exist between most health care markets and the theoretical model of competition. For the results to hold, several assumptions must apply.

The Assumptions under Perfect Competition

The First and Second Theorems apply to competitive markets. To be perfectly competitive, a market must have free entry and exit, perfect information, a homogeneous product, and numerous buyers and sellers each with no power over price. Furthermore, we derive the theoretical efficiency of competitive markets under conditions where no significant externalities, public goods, or natural monopolies exist. Finally, the actors in the competitive markets are alternatively consumers maximizing their utility, or producers maximizing their profits.

Many have criticized the application of the theorems to the health care sector, claiming that health care markets are typically not perfectly competitive. Health economists have recognized most of these criticisms as having validity. The health care markets depart from competition in several ways:

1 Barriers to entry exist in health care markets. Such barriers include licensure laws and health planning controls on prices and facility construction.
2 There are often few enough firms that those in the market have some degree of monopoly power.
3 Health care services are not uniform in quality or other characteristics.
4 Motivations other than pure profit are common in health care.
5 The model depicts the operation of markets under conditions of certainty. However, health events entail a considerable degree of uncertainty.
6 Information problems exist.
7 Externalities are prevalent in health care.

Several of the seven listed deviations need little further explanation. However, we consider three for extended discussion: the role of uncertainty, the role of information, and the role of externalities.

THE ROLE OF UNCERTAINTY The uncertain nature of health status gives rise to the demand for insurance coverage among persons who are risk-averse. In the present context, insurance creates problems for the efficient functioning of health care markets. We note four issues:

1 Insurance changes the price of care to the insured person, which in turn leads to the distortions described under price discrimination.
2 Insurance causes the price paid to suppliers to differ from the price paid by the consumer, and this distorts the efficient matching of production to consumption.
3 Large insurance companies and government programs negotiate payment rates, thus removing price determination, at least in part, from the market.
4 In some health care markets, insurance coverage is so complete as to distort the health care producer's incentives to be efficient.

THE ROLE OF INFORMATION The efficiency results for competitive markets depend on all parties having complete information available. As we have shown elsewhere, it is particularly problematic for markets to function when information is imperfect and asymmetrically available to the parties in the market. Potential problems of information and efficiency arise either when the physician has much more information about the appropriateness and effectiveness of treatments and techniques than does the consumer, or when the consumer has more knowledge of his or her health status and health habits than does the potential insurer.

THE ROLE OF EXTERNALITIES Finally, health care markets may involve externalities. A prominent externality will occur whenever participants in the market are significantly concerned about the health care received by others, not just about their own health care. This externality may be difficult to internalize in private charity markets, and it arguably causes health care markets to be inefficient. Because some analysts have identified externalities as the most important efficiency argument for social insurance programs in health care, we develop an extended discussion later in this chapter.

Promoting Competition in the Health Care Sector

If we could manipulate real-world markets as easily as we can change the assumptions of theory, then it would follow from our theoretical discussion that we should promote competition in health care markets whenever possible. Often the promotion of competitive elements in health care markets will prove useful. However, further theoretical grounds exist to qualify our statements.

The Theorem of the Second Best

One qualification involves the Theorem of the Second Best in welfare economics. Consider an economy with more than one departure from the conditions of perfect competition. Consider further any policy that corrects one or more of these departures from perfect competition but does not correct all of them. The Theorem of the Second Best shows that such a policy may not necessarily improve society's welfare.

An intuitive understanding of this result comes by considering a market with a pure monopolist (a departure from the conditions of perfect competition) who is also a polluter (a departure from the conditions under which competition is efficient). Basic theory shows that

a monopolist will produce less output than would a competitive industry under otherwise similar conditions. A policy that hypothetically converts the industry to perfect competition would resolve one discrepancy but not both because both output and pollution would increase. Societal valuations of the extra output versus the extra pollution could in principle determine whether the change worsened society's well-being. Thus, correcting some economic "wrongs," but not all of them, may not necessarily improve welfare.

This classic example of the monopolist polluter illustrates the idea of the theorem, but it does not make clear its applicability to the health sector. Consider a somewhat more controversial health-related example. Laws requiring that physicians go to medical school and pass additional exams grant licensed physicians a degree of monopoly power, a distortion from the competitive conditions. At the same time, however, health consumers have imperfect information on therapies and prices and less information than the physicians. This, too, is a departure from competitive conditions. If one eliminated the imperfection caused by licensure but did not simultaneously address the information problem, leaving patients less informed, patient welfare could decrease. This could happen because, without licensure restrictions on physicians, poorly informed consumers could be fooled by quacks or by possibly dangerous treatments. This example also illustrates the Theory of the Second Best.

It would be a misapplication of the Theorem of the Second Best to conclude that all health care policies that increase health care market competition are incorrect. More properly, the theorem states that we cannot assume competitive policies will always improve welfare. We necessarily operate in the world of second best because it will be impossible to convert all health care markets into the model of perfect competition. A competitive policy may improve the functioning of health care markets in a manner that improves society's well-being. Each policy must be considered on its own merits, not solely on the grounds that it promotes competition.

A number of health economists not only point out reasons why health care markets do not qualify as competitive, but they also criticize the very assumptions that underlie the efficiency claims of welfare economics. We will discuss these "extra-welfarist" claims in the section on Need and Need-Based Distributions later in this chapter.

An Economic Efficiency Rationale for Social Health Insurance

An externality occurs when someone external to the market transaction—that is, someone who is neither the buyer nor the seller—is affected directly by the transaction and does not receive compensation. A common example in health care occurs in the case of immunization for contagious diseases. Here, people outside the market transaction—those not presently immunized—benefit from the immunization because the immunized person will less likely become a carrier of the disease and threaten their health. This situation is an example of a beneficial consumption externality.

In the presence of a beneficial externality, the competitive market will tend to produce an inefficiently low level of output. Within a single market, the Pareto efficiency definition leads to the condition that marginal benefit equals the marginal cost in equilibrium. Individuals in a well-functioning, perfectly competitive market in theory will use medical care until the marginal benefits, measured through the demand curve, equal marginal costs, which in equilibrium will equal the price. In Figure 18.4, this leads to an efficient level of consumption, Q_m, in the absence of externalities.

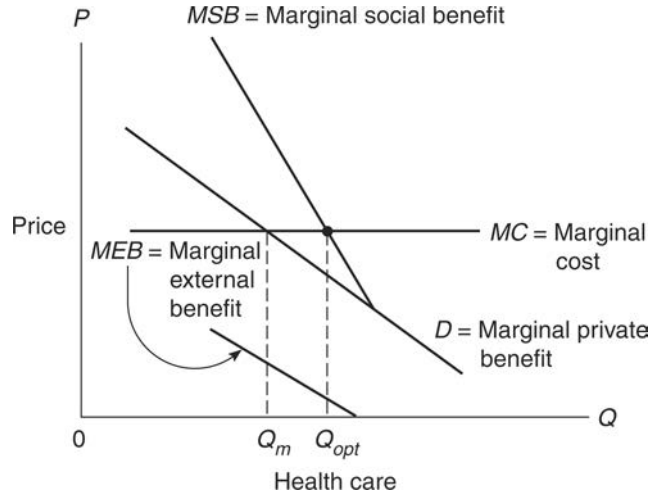


Figure 18.4 The Socially Efficient Equilibrium in the Presence of a Beneficial Externality

A marginal external benefit to people in society must be added to the marginal private benefit, which is measured by the demand curve, leading to the marginal social benefit. In Figure 18.4, the marginal external benefit curve is *MEB*. The marginal benefit to society as a whole is the vertical sum of the *MEB* curve and the demand curve. The result is the marginal social benefit curve, *MSB*. Efficiency for society occurs at output level Q_{opt}, whereas the market would achieve an inefficiently low level of output, Q_m. Thus, on efficiency grounds alone, society may be justified in subsidizing immunizations.

Although immunization for contagious diseases illustrates the logic and role of beneficial consumption externalities in justifying subsidies for (or possibly public provision of) care, such as the U.S. polio immunizations of the 1950s and 1960s or modern-day immunizations in less-developed countries, it represents a fairly minor problem and could not in itself be used to justify large social insurance programs. However, an alternative health care externality, one that we have identified elsewhere as a charitable externality, can in principle be sufficiently important to justify such programs.

This externality would occur, for example, whenever people feel that some segment of society is receiving insufficient care in the sense that the charitably minded person would be willing to pay to help these people get care. Willingness to pay means here that they would pay if contribution would help the poor to acquire health care. Such charitable feelings are probably widespread in most societies. As Pauly (1971) argued:

> The desire to eliminate the diseconomy that the presence of curable but uncured disease or injury may exert on others does appear, in general terms, to be a common characteristic of human beings. At least at some levels, most of us would be willing to give up some of our income to help a suffering fellow. Some may, of course, be immune to such feelings, but individuals may also be immune to contagious disease, and this should cause no insurmountable theoretical problems.
>
> (pp. 10–11)

This charitable externality has a different source than the case of immunization and is probably more important, but the economic argument is of the same form. In Figure 18.4, consider the curve *MEB* to measure this externality. Then, as before, the efficient level of output, Q_{opt}, exceeds the market output, Q_m. Under certain circumstances, this efficiency may be sufficient grounds to intervene, for example, by providing a program of social insurance.

Need and Need-Based Distributions

Even if we can identify accurately the efficient allocation of health care, we nevertheless may find many people in society dissatisfied with the outcome because many people will not get the health care they need. That is, there will be additional concerns, over and above efficiency concerns, regarding equity. In the health care literature, the concerns for equity most often center on the question of whether people are getting the health care they need. Unfortunately health care need is often either undefined or variously defined. The definitions employed may take either extreme, either maximally or minimally. For example, Culyer and Wagstaff (1993) review several definitions of health care need; they conclude that a maximal need definition is superior, defining need as:

> the expenditure required to effect the maximum possible health improvement or, equivalently, the expenditure required to reduce the individual's capacity to benefit to zero. (p. 436)

In contrast, some analysts or policymakers treat health care need as a minimal requirement or standard of adequacy. Federal health planning efforts in the 1970s sought to control the perceived proliferation of health care in order to control costs. Health planners at the time, as well as most health cost control advocates in any era, argued implicitly that consumers were getting more health care than they really needed.

Often the discussion of needs gets disconnected from the fact that the output and distribution of health care to meet people's needs are chosen in the context of society's choices of all its public goals. We present a construction of need that brings this to mind. Let health care needs be defined within the context of the choice of society's goals for population health status as well as in the contexts of other goals, such as education and defense. To illustrate, we identify in Figure 18.5 the production function for health defined over the levels of a variable input, health care, given the conditions of environment *E*, lifestyle *LS*, and human biological endowment *HB*. Here the technically maximal health status is HS_{max}. Achieving this health status level requires a health care level of HC_{max}. However, society may choose a lesser health status goal than the maximum achievable, using the savings to further other goals. For example, if society through its choice-processes selects health status goal HS_0, then health care level HC_0 is needed.[2]

The choice of a health goal implies a needed level of health care, that is, a level of health care "utilization." In the early 1990s, economists debated whether utilization or "access" was the superior choice for defining needs. The words *equity of access*, which frequently appear in public documents in various countries, suggest an equal opportunity, especially a financial opportunity (Mooney et al., 1992). Many health economists, however, find *access* difficult to define. While health care utilization is more easily measured (in units such as visits, or days of care, or availability of necessary drugs), it is usually also the ultimate reason for our concern about access (Culyer, van Doorslaer, and Wagstaff, 1992a, 1992b).

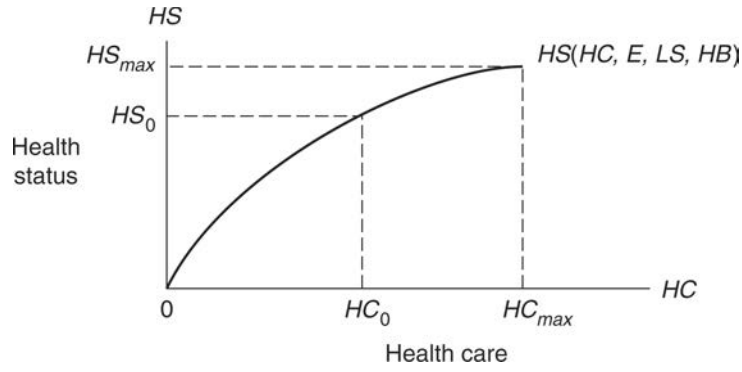


Figure 18.5 The Production Function for Health

Health Care Needs and the Social Welfare Function

How, then, is the health status goal selected? We depict this choice using the concept of a social welfare function.

THE UTILITY-POSSIBILITY FRONTIER The Edgeworth box shows the efficient choices available to society in allocating resources among people. Yet it is also apparent that at many points on the contract curve, Abner or Belinda get few or no resources. By the Pareto criteria, these points are unquestionably economically efficient, yet they may be indefensible within any definition of a humane society.

Economic theory suggests that a social welfare function reflecting society's overall preferences is necessary to determine which of the efficient points to choose. We can trace out a utility-possibility frontier, *UU*, from the points in the Edgeworth box in Figure 18.1. Begin at the allocation where Belinda has everything and Abner has nothing. In Figure 18.6, we can assign Abner a utility of zero and evaluate Belinda's utility as the intercept of the vertical axis.

We can then draw a *UU* curve by reallocating resources to Abner from Belinda as we move along the contract curve in Figure 18.1. Recall that the fundamental property of Pareto efficient distributions is that Abner's utility can come only at the expense of Belinda's utility. Therefore, the *UU* curve in Figure 18.6 must be downward sloping. The horizontal axis intercept summarizes the point at which Abner has all of the resources.

THE SOCIAL WELFARE FUNCTION Society must now decide which point on *UU* to choose according to the rules by which societies operate—through debate, consensus, and maybe even dictatorial power. Economists define this set of rules as a social welfare function. Consider an unusual society in which the consensus was that everyone's utility level should be exactly the same, with no variations tolerated. We would recognize this as a set of right-angled social indifference curves along a 45-degree line from the origin. The optimum allocation would be at point *A*, which is a tangency between the *WW* curve and the *UU* curve. This would indicate equal utility levels for Abner and Belinda. From point *A*, we can then return to the one point on the contract curve in Figure 18.1 in which the utility levels are equal, denoted *Z*. Selecting this point leads to the unique allocation of the two goods to Abner and Belinda (although not necessarily the same amounts to each, as shown in Figure 18.1).

Many would find the specific social welfare function with equal levels of utility highly questionable and almost impossible to define or obtain (realizing this, policymakers often seek to redistribute the goods that provide the utility). Humane societies might agree, however, that

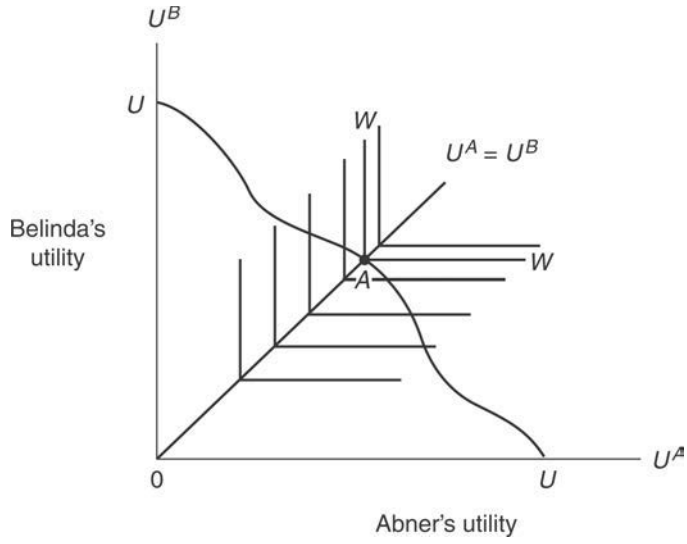


Figure 18.6 The Utility-Possibility Frontier and Social Welfare Maximization: Egalitarian Preferences

everyone should receive enough resources for at least a minimal standard of living: for Abner this would be U^A_{min}, and for Belinda this would be U^B_{min}.

We can characterize this condition as two constraints on the *UU* curve in Figure 18.7. Such societies would consider no social welfare function that would provide Abner with less utility than U^A_{min}; this would be similar for Belinda regarding U^B_{min}. Even here, however, the answer depends on the exact social welfare function chosen. Societies in which people like Belinda are most influential may have social welfare functions with social indifference curves like *W′W′*. In this case, people like Belinda would get larger shares of the resources than people like Abner. The optimum at point *B*, although providing subsistence living for citizens like Abner, would leave citizens like Belinda better off than the optimum in Figure 18.6.

THE SOCIAL WELFARE FUNCTION AND HEALTH CARE NEEDS Within this framework, we now discuss the various social choice processes that scholars and policymakers have proposed for the equitable provision of health care. Let the social welfare function of society (*SW*) represent the preferences of society as a whole. The function in a commonly used form is:

$$SW = f(U_1, U_2 \ldots, U_n) \tag{18.1}$$

where social welfare is characterized as a function of the utility levels of each of the *n* persons in the society. Utility for each person as usual depends on his or her consumption of the available goods in society, including health care. In a variation, introducing caring for others, we may suppose that each individual to some degree perceives an external benefit from the consumption of health care by others.

We may perceive society to be efficient when it acts as if it were choosing among its variables to maximize the social welfare function. Consider the choice process graphically in Figure 18.7 as choosing the highest social indifference curve attainable given the utility-possibility frontier. This leads to the appropriate level of health care. The health care needed by each person in society is that level which maximizes *SW*. Social welfare will be maximized when society chooses its optimal health status goals in conjunction with optimal levels of other goals.

The social welfare function has proved flexible within health economics to formulate other conceptions of health equity. Yet a lot of what economists do when they are advising the public lies outside standard welfare economics. In Box 18.1, we have suggested several "extra-welfarist" criticisms. It may now be beneficial to define and contrast the relevant terms more thoroughly.

BOX 18.1

The Extra-Welfarist Critique

Werner Brouwer and colleagues (2008) do an excellent job of drawing distinctions between welfare economics and extra-welfarist views, and we consider their work here. Some extra-welfarist criticisms go well beyond the usual criticisms of economics, directed more at the basis of welfare economics itself:

1 Consumers may not be rational.
2 Individuals may not be the best judges of their own well-being.
3 Social welfare may depend on more than individual utilities.
4 Consumer tastes are not fixed but are often learned and malleable.

These issues, recognizable to economists, raise serious objections for the analysis we have described. For example, if the assumption of rationality fails to approximate behavior, then most microeconomic theory would need to be re-evaluated.

While items 2 and 3 further show how challenging the normative issues are, they are not foreign to America. U.S. society often makes choices that imply that individuals are not the best judges of their own well-being. Examples include motorcycle helmet laws, criminalization of drugs, and mandatory retirement contributions.

However, controversy always arises over where to draw the line. In applying welfare economics to the uninsured, Mark Pauly suggests that voters could probably be convinced that the value of certain reforms aimed at reducing the rate of uninsurance is worth the costs. However, he asserts, "If we cannot convince the decisive voters of the value of what we value, then I think we need to accept the verdict of democracy" (p. 14). Uwe Reinhardt responds in contrast:

> I, for one, believe that, if this nation is ever to have truly universal health insurance coverage and a truly humane safety net all around, an elite espousing those goals would have to impose that state of affairs on generally confused plebs that has quite unstable, often logically inconsistent and utterly malleable preferences on the matter.
>
> (Pauly and Reinhardt, 1996, p. 24)

Reinhardt echoes the frustration of many health economists on this point, but many also question it. *Plebs* here means the general populace, but it shares the same root with the word *plebiscite*, a vote of the general populace. Would health policy "impose(d) . . . on a generally confused plebs" pass a vote, or plebiscite?

WELFARIST HEALTH ECONOMICS Thus the social welfare function (SWF), which we have used to choose the "best" social option in Figure 18.7, though the dominant approach, is disputed by many health economists. Called "welfarism" by its critics it is contrasted with "extra-welfarism." It will be beneficial to define and describe it more thoroughly here.

The SWF represents welfarism when based solely on the utilities of the individuals who make up society. As in much of economics, we assume that these individuals are rational and that they know what is best for themselves. Everyone counts. If John would improve his utility when society moves from *A* to *B*, and if no one else is harmed, then society must choose *B*. Thus the Pareto Principle, which you recognize here, applies in the SWF.

This welfare economics contains two other salient features that matter when we compare the extra-welfarist argument. First, note that the mathematical form of the SWF and any equity weights that may be incorporated in it come from "outside" of the assumptions described previously. Perhaps they come from a societal advisor, or, more in keeping with the individualistic and democratic nature of the approach, perhaps society could determine preferences over these features by surveys of the public.

Second, the concept of utility is in some places variously defined. The introductory classroom will find it defined as "a measure of satisfaction" or even "happiness." The more sophisticated definition describes utility as merely an index of preferences; where preferred

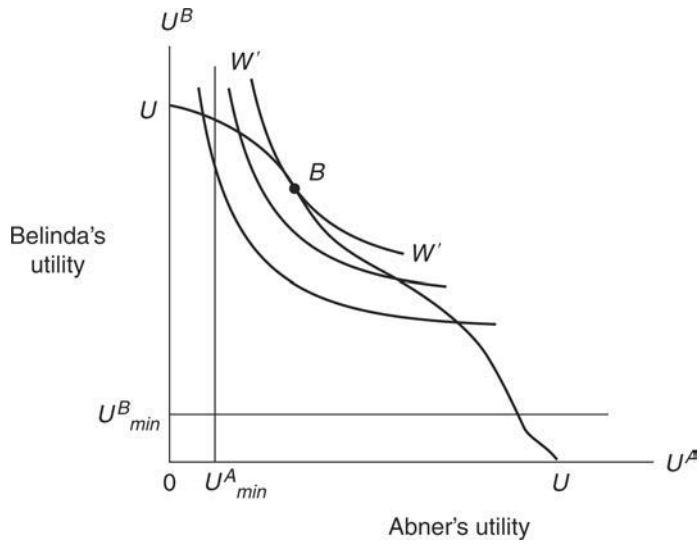


Figure 18.7 The Utility Frontier and Social Welfare Maximization: Preferences Favoring Belinda

bundles get higher utility numbers. The latter definition puts an emphasis not on emotion but on what the individual would choose under the right conditions.

EXTRA-WELFARIST ANALYSIS In contrast, the extra-welfarist approach allows valuations other than the individual's utility in the determination of social choice, and the sources of valuation may be other than the affected individual's. Therefore it is less individualistic and tends to be less democratic, at least in the sense that unlike the "methodological individualism" of standard welfare economics, extra-welfarism makes no assumption that the individuals' preferences are the best guide to what is best for themselves.

Sen (1980), for example, questions whether a person's utility is a reliable guide to his or her well-being. He argues that utility may merely result from, or represent, the emotions of the moment. He proposes, instead, that individuals are entitled to an acceptable level of "capability," which includes health and normal functioning. Ruger (2009) elaborates on Sen's important arguments, exploring particular policy implications. Cookson (2005) has proposed that the quality of life-years (QALYs) measure provides an acceptable approximation to Sen's capabilities. An interesting twist of this theme is Alan Williams's Fair Innings approach, described in Box 18.2.

BOX 18.2

The "Fair Innings" Proposal

Williams (1997) proposes that people would generally agree to the principle that everyone is entitled to a normal span of life at a reasonable level of quality. The Williams metaphor fits on both sides of the Atlantic: Cricket has one or two innings and baseball has nine innings—generally that is all anyone gets—but everyone should get that much. This idea applies with most force to trade-offs in life-years between the old and the young; in this context, it implies a rationing by age.

Consider two individuals, each of whom stands to gain eight good-quality extra years of life after being "rescued" from a lethal disease by medical technology costing $100,000. Let one of these people be 80 years of age and the other 30. If resources are scarce, which individual should get the treatment? The "fair innings" concept would clearly require the treatment go to the younger person since the 80-year-old person has had their fair innings already.

Valuations by society in this fair innings scheme are generally not so simple. This central issue lies in how much society prefers to help the one versus the other. If we agree that the elderly facing illness and the young facing illness are not equal, then the question for Williams becomes "To what degree are we as a community averse to this inequality?"

Norman Daniels's Concept of Health Care Need

The social welfare function framework implicitly treats health care as just one of many commodities that provide utility, something determined by the same social choice mechanism by which society makes all its choices of social goals. Some analysts argue instead that health is special and that health care needs have a more objective and independent basis.

To address the special role of health, the philosopher Daniels (1985) developed a health care needs definition, and we outline his approach here.

1 ***Health care is special.*** This view separates health care from other goals. To make this separation plausible, we must view health care as a primary good that is special because it is central to the task of attaining or restoring a fair equality of opportunity. In this sense, it comes prior to other considerations.
2 ***Species-typical functioning.*** Daniels argues further that the human species has a range of functioning that is typical and appropriate to it. *Disease* here means the absence of health; *health* is the ability to attain a functioning level typical to the species.
3 ***Fair equality of opportunity.*** Given the nature of society and the human species, there is a range of behavior opportunity that every person in society should have available. The range will vary somewhat from person to person inasmuch as we each have different endowments of skills and abilities, but all are entitled to their fair share.

To use the fair equality of opportunity standard of health care need, we must imagine that there is a degree of objectivity behind its construction, and that need, so constructed, would be observable in common by most people. We do not solve this matter here, but instead point out what Daniels's project entails. If we accept Daniels's view as correct, then health care need is separate from the other goals for society. We leave as a discussion question the issue of whether this is the case.[3]

Economic Criticisms of Need-Based Distributions

Mainstream economic ideas often clash with need-based conceptions of the appropriate distribution of health care resources. The economic criticisms are directed to particular conceptions of need and can be understood as saying "If by need you mean X, then the following criticism applies. . . ." Several criticisms of this sort exist in the literature.

1 ***The bottomless pit.*** If the technical maximum health status goal is chosen, it will likely be the case that the cost of bringing all inputs to bear for some patients, even when these inputs have little effect on the patient's health, could exhaust society's resources. This would constitute a bottomless pit.
2 ***Needs should not be chosen independent of costs.*** Society's health status goals should depend in part on the costs of health and the price of health care. Health is undoubtedly subject to diminishing marginal returns in production. If some other societal goal offers greater marginal utility per dollar than health care, society could improve its well-being by transferring the money value of the marginal health care unit to the service of that other good, for example, education.
3 ***The role of scientific medicine in determining needs.*** From similar reasoning, it follows that health needs cannot be determined solely on the basis of scientific medical knowledge. The role of medical experts is critical in needs analysis, inasmuch as we require scientific data to determine the medical inputs needed to attain a given health goal. However, the appropriate health status goals themselves must be chosen with knowledge of society's economic constraints and its values. In some form, the political process is required to identify the trade-offs that society is willing to incur to attain any given goal.
4 ***Monotechnic needs.*** Early on, Fuchs (1974) pointed out that when needs analyses are stated in terms of needed health resources per capita, they often implicitly (and usually incorrectly) assume that only one available technique exists for pursuing a health status goal. More plausibly, many opportunities exist for substitution, not only among health care inputs but also between health care inputs (as a group) and other inputs such as improved diet or exercise.

Horizontal Equity and Need

Horizontal equity is the requirement that equal people be treated equally. Analysts have often compared health care equity across countries using a Gini Index, or at least one modified from the original Gini's use, though recently researchers have developed measures with more attractive features. Consider first what the traditional Gini Index is.

Figure 18.8 depicts the cumulative portion of the population ranked by income (on the horizontal) graphed with the cumulative portion of earned income on the vertical. The diagonal line indicates the "perfect distribution." Along this line each income group is earning an equal portion of the income. The broken curve line represents the Lorenz Curve, the actual situation. For this example, the bottom 20 percent of the income distribution is earning 5 percent of the income. Income inequality favoring the rich occurs whenever the actual curve lies below the diagonal. The Gini Index is the area marked *A*, and the Gini Coefficient is the ratio of area *A* to the total area under the diagonal.

We see that by these definitions the Gini Coefficient must always lie between zero (perfect equality) and 1 (complete inequality); in other words, the Lorenz Curve must always lie below the diagonal. For example, it would be illogical to say that the lowest 20 percent of the income people received more than 20 percent of the income. But things are different when health care inequality is the issue.

Suppose we wish to compare the cumulative proportion of health care use with the cumulative distribution of income. In contrast to income, it is not illogical to say that the lowest 20 percent of the income distribution receive *more* than 20 percent of the health care. Such a distribution would be "biased" in favor of the poor. We shall see that the distribution of health care favors the lower income groups in many of the European countries. We can imagine in such a case a figure somewhat like Figure 18.8 except that the actual distribution curve has portions that lie above the diagonal. By convention we associate bias toward the poor with negative numbers and bias toward the rich with positive numbers.

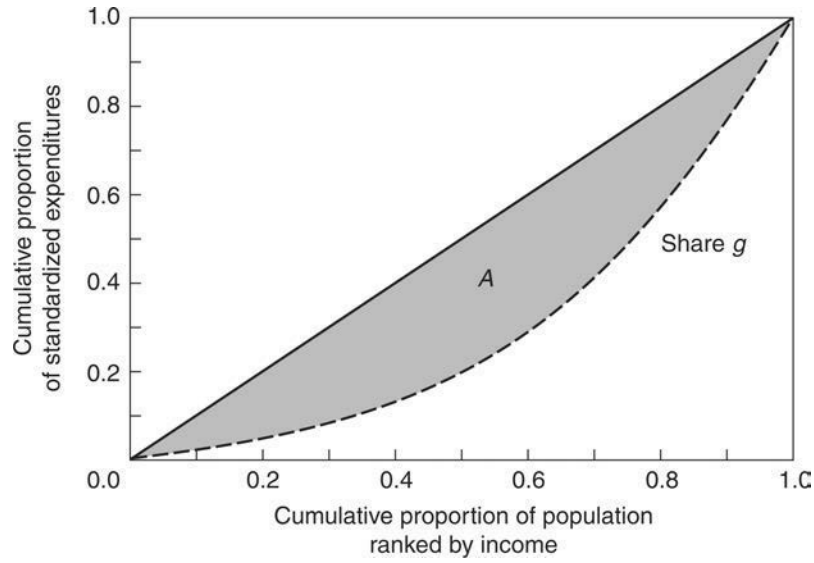


Figure 18.8 The Gini Coefficient

While the Gini analysis still embodies the basic ideas of horizontal equity analysis, health economists (van Doorslaer, Koolman, and Jones, 2004; Koolman and van Doorslaer, 2004) have developed further measures and extensions of the analysis, which they find more useful. These authors prefer a Concentration Index, stated as

$$C_M = \frac{2}{\bar{y}} \operatorname{cov}(y_i, R_i) \tag{18.2}$$

where y_i is the health care utilization of income group i, $\bar{y}$ is the mean health care use in the population, and R_i is the cumulative fraction of the population in fractional income group i. Equation (18.2) states the unweighted version of the index, which is correct when all of the groups are of equal size. If the groups have different sizes, a weighted version must be used, as in van Doorslaer, Koolman, and Jones (2004). Under the present assumptions, the unweighted covariance of y_i and R_i is

$$\operatorname{cov}(y_i, R_i) = \sum_{i=1}^{i=n} \frac{(y_i - \bar{y})(R_i - \bar{R})}{n} \tag{18.3}$$

Equation (18.2) shows that a bias favoring the rich will yield a positive covariance and thus a positive value for C_M. Consider this statement in more detail. Note that each term under the summation sign in (18.3) is a product of the individual's relative amount of care times his or her relative position in the income scale. If richer than average, $\left(R_i - \bar{R}\right) > 0$, and if at the same time receiving more care than average, $\left(y_i - \bar{y}\right) > 0$, then that product will be positive. Similarly, if poor and receiving less than average care, the corresponding product would also be positive.[4]

If the predominance of care likewise tends to favor the rich and disfavor the poor, the covariance will tend to be positive. Conversely, a bias in favor of the poor will tend to result in a negative covariance. Returning to (18.2), we understand that a positive value for C_M suggests a bias in favor of the rich and a negative value for C_M suggests a bias for the poor.

In the data that follow, the authors have created the Concentration Index for several European countries, and they have also created a Health Inequity (*HI*) index. The *HI* is calculated by first creating a Concentration Index for health *need*, denoted C_N. They then subtract $C_M - C_N$ to control for the need-based variation. The reason behind this step, in the authors' view, is that we should not view health care visits to the doctor that respond to immediate need as treatment caused by inequity of the health system itself.

To devise their estimate of health need, the authors regressed doctor visits on "need indicators." For this purpose they chose health status measures, morbidity (illness), and demographics. They then created C_N by replacing the values for y in equation (18.2) with this measure of an individual's health care need.

The Health Inequity index is

$$HI = C_M - C_N \tag{18.4}$$

Both indexes are reported in Table 18.1. Index C_M is the Concentration Index as we have previously described it. Index *HI* may be thought of as health care inequality after removing the underlying variation that can be attributed to need. Note that C_M and *HI* estimates in the table are predominantly negative for these countries for General Practitioner visits, indicating a bias in favor of the poor.

However, for the specialty visits the reverse is true almost throughout. Table 18.1 gives us two means to assess the degree of bias in these countries. First, the sign of the indices indicates a bias toward the rich if the index is positive and toward the poor if the index is

Table 18.1 Health Care Inequality Measures across Several Countries

	C_M	*HI*	C_M	*HI*
	GP Visits Total	*GP Visits Total*	*Spec Visits Total*	*Spec Visits Total*
Ireland	–0.1323*	–0.0696*	0.0770*	0.1388*
Belgium	–0.1145*	–0.0508*	–0.0269	0.0255
Spain	–0.0906*	–0.0492*	0.0267	0.0740*
Luxembourg	–0.0918*	–0.0406*	–0.0704*	–0.0282
Italy	–0.0649*	–0.0349*	0.0179	0.0537*
Greece	–0.1258*	–0.0308*	–0.0418*	0.0492*
Germany	–0.0636*	–0.0268*	0.0158	0.0517*
U.K.	–0.1006*	–0.0240*	–0.0234	0.0524*
Netherlands	–0.0535*	–0.0113	–0.0178	0.0413*
Denmark	–0.0831*	–0.0008	0.0223	0.0844*
Portugal	–0.0692*	–0.0051	0.0971*	0.1604*
Austria	–0.0499*	–0.0146	0.0345	0.0740*

Notes: C_M is the Concentration Index and *HI* is the Health Inequality Index, both of which are described in the text. The table is created from data published in van Doorslaer, Koolman, and Jones, *Health Economics*, 2004, Tables 1 and 2, pp. 637–38, with permission. * Indicates that the estimated value is significant at the 5 percent level or better. "GP" stands for general practitioner, and "Spec" stands for specialist.

negative. Second, we measure the strength of the bias by the size of the index and whether the index estimate differs significantly from zero at the 5 percent level.

Income Inequality

To have lesser access to health care raises the concern that health status will also be lessened. Income inequality has a broader effect on health. We show the relation of income inequality to mean health status by defining it as the mean infant mortality rate for two reasons: (1) deaths can be counted with little error; and (2) most countries and most smaller jurisdictions (like states) keep track of these data.

In examining Figure 18.9, we must be aware that access to health care is only one part of determining health status, whereas income inequality may have many additional effects, often harmful: for example, pockets of poverty may lack good nutrition. Figure 18.9, from the *Spirit Level* by Wilkinson and Pickett (2011), will surprise many American students and other readers who expected us to be doing much better than this. More important, the international pattern is clear and worth committing to memory. We see that income inequality, which economists usually measure by the Gini Index, tells us much more than mere per capita income data. Countries that are rich on average, like the United States, may be less healthy than many others who share their wealth better.

One may have thought that the Gini and infant mortality would be unrelated, but income equality proves to be closely related to this and much more. Again following Wilkinson and Pickett (2009): child well-being (UNICEF Index) improves with greater equality; it improves

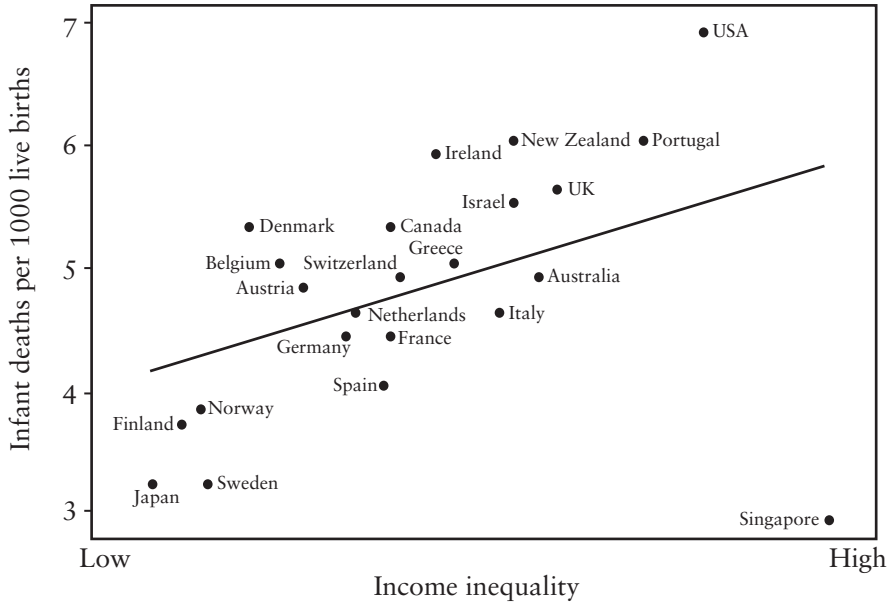


Figure 18.9 Infant Mortality Is Related to Inequality in Rich Countries
Source: Richard Wilkinson and Kate Pickett, *The Spirit Level: Why Greater Equality Makes Societies Stronger*, New York: Bloomsbury Press, 2009.

"trust" between people in the community; women's status improves; mental health problems are lower; life expectancy is extended; children get greater math and literacy scores; there are fewer prisoners per capita.

The influence of income inequality is clearly substantial. But we know that correlation is not causality. Could it be that these data actually reflect that those people who are healthier to begin with will more likely generate income equality? We do not have scientific papers to resolve this question (though see related discussions on social capital in Chapter 24). Yet many if not most economists would agree that reducing income inequality would likely improve community well-being.

Schooling and Income Inequality

Political scientist Robert Putnam provides a needed micro-level look at the effects of income inequality (2015). He and his staff talked to selected parents and their children in several American towns and also studied the local schools.

- Clara first lived in a poor and dangerous area of L.A. But with encouragement from her teachers and very determined study she qualified for college and went on to get a graduate degree in counselling. When she married Rodrico, also a professional, they had sufficient means to care for their daughter, Isabella, and move to an affluent area of Orange County near an excellent high school, Troy. Isabella attends Troy.
- Lola and Sophia don't remember their mother, and they live in a house, provided for by their grandfather, who lives elsewhere, in a very bad part of Orange County, just a few minutes' drive from where Isabella lives. They attend Santa Anna High School, which the

girls complain to be undesirable. Ironically the school receives about the same amount of public support per student as does Troy, though Troy receives donation support from its wealthier parents.

Troy students get much higher mean SATs. Does this academic quality matter to a given student's (e.g., Isabella) achievement?

The answer is "Yes." Research (Oldfield and Eaton, 1996) shows that characteristics of your school peers, whether measured by test scores or family income, are more strongly correlated to one's own achievement than one's own characteristics. Others have shown that income class of the local families is a key determinant of student achievement in the local schools (Reardon, 2011). So what has income inequality to do with this situation? First, when families move freely, they will buy a house, a neighborhood with pleasant amenities, and quality schools. The key word is "buy"; even the neighborhood amenities have an effective price on them. Sophia and Lola had little or no wherewithal to get any better situation. Isabella's good education is provided by the investments her parents had made in Los Angeles.

Second, families that provide stimuli for learning can have substantial effects on their children's academic success: reading to them, talking with them, encouraging digital literacy, less time watching TV, and more. Does this academic richness at home have consequences? Yes. Research reported by Nisbett (2009) shows that it will even increase the children's IQ scores.

Can public policy improve the lot of poor families? Recall that Clara succeeded by determination. Of Lola and Sophia, Lola dropped out of school while Sophia determinedly fought the lax system at Santa Anna and eventually qualified for Community College.

Is determination enough, and can academic achievement be left to the children themselves? Research by Kenworthy (2012) finds that simply giving a quantity of dollars (he used $3,000) to a poor family with young children will benefit the children now as well as later in life.

Nobel economist Joseph Stiglitz (2012) sets out a collection of reforms to improve income equality in the United States. We list a sample of these so readers can assess their logic, but also to show how politically difficult the path would be: (1) make income taxes more progressive; (2) end corporate welfare; (3) more effective enforcement of competition laws; (4) and reduce rent seeking.

Theories of Social Justice

Inevitably, understanding what health care distribution is equitable and choosing what health care needs should be met in a society depends on ethical theory. An ethical theory serves to identify a context and reasoning by which to determine what we *ought* to do, as opposed to mere positive analysis which describes what we do. Ethical theories that serve to determine a fair or just distribution of economic resources are sometimes called theories of social justice. Seen this way, any notion of equity or need in health care, to be complete, must be connected to an ethical theory, perhaps to a theory of social justice.

Although there are several theories of social justice, there is no consensus-accepted theory. Even without a consensus, however, such theories help to illuminate issues to address in order to achieve a consensus. We offer a brief overview of three social justice theories, along with a selection of criticisms of each theory.

Utilitarianism

Utilitarianism became prominent in the nineteenth century and is still current in modified forms. It can be understood as the greatest good for the greatest number. In its classical form,

it identified the social optimum coincident with maximization of the sum of utilities of all persons in society. Classical utilitarianism essentially defines the social welfare function as the sum of individual utilities.

Followers of utilitarianism, promoted by Jeremy Bentham in the eighteenth and nineteenth centuries, conceived of an individual's utility as cardinally measurable, at least in principle, and comparable among individuals (your "utils" compare with my "utils"!). The utilitarian ethic originally was conceived somewhat literally as maximization of society's total satisfaction level.

Utilitarianism captures the idea of trade-offs among goals. Under this construction, society may choose to accept some harm for a few members in return for a greater good for the many. As such, it avoids the bottomless pit criticisms mentioned earlier. Health status would not generally be maximized for every individual in society under this view.

SOME CRITICISMS OF UTILITARIANISM Economists came to criticize classical utilitarianism early in the twentieth century. They generally rejected the idea that utility could be cardinally measurable and comparable among people. Economists view it as unscientific to suppose that one individual's level of satisfaction could somehow be added to that of another person. Modern social welfare theory in economics has proceeded along ordinal utility lines. Utility in these theories retains the role of ranking preferences among alternatives, but we have discarded the notion of a fixed quantitative measure of happiness in most modern theory.

Two other criticisms of utilitarianism illustrate some of the weaknesses identified in the theory. One is the question of domain—that is, whose utilities are to count? Utilitarianism does not itself identify where to draw the boundaries of membership in the society. Are foreign people or noncitizens to count? If not, why not? Are animals to count? Unborn future generations? Is the utility of the fetus to count or only that of the already born?

A second criticism raised by Nozick (1974) poses a similar question regarding possibly malevolent individuals in society. For example, suppose an individual, because of bigotry or sheer malevolence, gets satisfaction out of the suffering of some other group in society. Is the malevolent utility of such a person also to count?

Rawls and Justice as Fairness

John Rawls (1971) approached the concept of social justice from a different viewpoint. Here, a primary principle of justice is that social choices must be fair. Rawls views it as unfair for people with economic or political power, who often have vested interests because of their circumstances in society, to dominate the social choices. Instead, according to Rawls, to be fair we should make our choices from a position divorced from arbitrary special interestedness. Such a position, it is proposed, is one from behind the "veil of ignorance."

The Rawlsian veil of ignorance is a hypothetical situation in which we can think rationally but for which our particularities of self and economic situation are as yet unknown. It is as if we could somehow contemplate life in society before we are born and before we know whether we will be rich or poor, black or white, male or female, tall or short, and so on. Rawls believes that, so divorced from our vested interests of life, we would generally come to a consensus about principles of social justice, and specifically we would agree to the Rawlsian "maximin" principle.

Under the maximin principle, we would each reason that without knowing who we were to become in society, we would presume that we could be the person worst off. Under such circumstances, we would agree, argues Rawls, only to a system of justice in society that maximized the position of the worst off. This need not result in complete equality of incomes and

resources including health care, but it would permit departures from equality only if the lot of the worst-off would improve. Health care under a Rawlsian system of social justice would presumably also be provided if the needs of the worst-off were regarded as a priority.

SOME CRITICISMS OF RAWLSIAN JUSTICE Rawls's theory of justice also has drawn criticism. Critics have noted that Rawls assumes that each of us behind the veil of ignorance is extremely risk-averse. Suppose that an alternative situation, A, offered everyone an income of $10,000, while alternative B offered one person $9,000 but everyone else $100,000. Under the maximin principle, persons behind the veil would choose alternative A, the alternative with the higher income for the worst off. Would people really be so risk-averse as to forgo even extremely good odds of a large gain? The Rawlsian theory of justice also appears subject to the bottomless pit argument. The instance of health care provides a good example of the problem in the views of Arrow (1973):

> Thus there could easily exist medical procedures which serve to keep people barely alive but with little satisfaction and which are yet so expensive as to reduce the rest of the population to poverty.
>
> (p. 251)

Despite the criticisms, Rawlsian justice provides a prominent example of a theory of social justice that entails a strong presumption in favor of equality—a presumption that permits inequalities to arise only if they contribute to the lot of the worst-off.

Liberalism, Classical, and Modern

Classical liberalism refers to the political philosophy developed largely during the Age of Enlightenment, which centers on the eighteenth century. Led by the principles of John Locke (1690), it emphasizes the rights of the individual to his property and to himself. In this view, people enter the state voluntarily, and are free to choose what they deem best for themselves and their families. The tradition was the intellectual guide to the American Constitution, as is well-known to American schoolchildren. These "classical liberals" also included Adam Smith and, later, John Stuart Mill and Friedrich Hayek. Liberty in this tradition was largely a matter of emphasis as opposed to a fixed constraint. These philosophers largely supported taxation and often mentioned favorably things, such as government programs, to improve the well-being of the community; Locke (1697, p. 4), for example, favored government aid to the poor by way of the English Poor Laws, Smith favored public schooling for working-class children (1776, pp. III, II, 303), Mill mixed liberty values with social concerns (1975), and Hayek, whose life spanned the creation of the New Deal, spoke favorably of social health insurance itself (1960, p. 298). Would classical liberals, in sum, support, for example, a modern universal social health insurance plan? Folland (2005) reasons that they might do so.

In contrast to the older, verbal tradition of these economic philosophers, Nozick (1974) departs both in analytical style and by drawing stronger implications. Nozick asserts that government must limit itself to maintaining only the necessary services: the "minimal state."

Nozick proposes that natural rights suggest the necessity for a *libertarian constraint*. By a libertarian constraint, he means that any system of social organization should prohibit the coercion of others, and that people are entitled to keep any property received through a voluntary transaction. From these principles, Nozick justifies the existence of a minimal state. He argues that these principles of justice necessarily limit the role of the state, and in consequence, this would exclude social programs beyond the minimal functions of government in providing public police protection services. It follows that social programs providing for health care also

would not be warranted, and the health care needs of one person would not place obligations on any other person other than for what he or she voluntarily is willing to accept.

SOME CRITICISMS OF THE MODERN LIBERTARIAN THEORY A central focus of criticism is the assumed libertarian constraint itself. For those who do not accept the constraint as an implication or necessity of natural rights, it will be hard to see why we are not free to trade off some degree of liberty in order to make gains in efficiency. For example, we do not allow slavery even if the would-be slave would voluntarily accept it. In the health setting, the Food and Drug Administration restrictions on the availability of certain drugs represent restrictions on liberty, but at least, in some cases, they may be supportable on efficiency grounds. Pauly (1978) has argued that if the costs of conveying information to the public are sufficiently large, substantial efficiency gains may accrue simply by restricting access to a potentially dangerous drug. The restriction of access may in some cases require a physician's prescription; in other cases, it may require banning the drug from the market.

Members of society frequently accept rules that restrict liberty somewhat but that are expected to improve outcomes. Wittman (1982) has discussed the potential efficiency of simple rules in day-to-day life and in sports. A traffic light restricts liberty but promotes the efficient and safe flow of traffic. If we accept the principle that liberties can be traded off to gain certain efficiencies, this raises the question of whether we must accept the libertarian constraint.

Conclusions

The three theories of social justice described here attempt to ground our understanding of the distribution of goods and services—including health care—in a system of ethics. The brief review of these theories cannot account for all arguments and rebuttals nor does it constitute an exhaustive coverage of the many normative models available. Those interested in ethics, as it concerns economic distribution, should consult the original sources.

The theories, however, serve to show how widely people's understanding of appropriate distributions of health care differ, and their consideration suggests that we have as yet no consensus. Nevertheless, these issues of justice are raised whenever society wishes to modify the distribution of health resources on grounds of need and equity. The issues are no less important because there is disagreement.

The meaning of efficiency is more sharply defined. Here the controversial issues involve the degree of efficiency attainable either with existing health care markets or with health care markets as modified by new policies. The theoretical model of perfect competition generates a Pareto efficient outcome. Health care markets in practice, however, differ in many respects from those that have perfect competition. Perhaps the most notable discrepancies of actual health care markets from the theoretical model arise because of the role of uncertainty, the problems of information, and the presence of externalities.

Perfection in either direction is not attainable. We must inevitably accept approaches to health care distribution that are second-best, evaluating each proposal on its merits. While perfect competition is unattainable, proposals that improve the degree of competition in health care markets may nevertheless improve society's well-being. On the other hand, proposals for providing social insurance programs to at least some segments of the population can be supported in principle on efficiency grounds whenever substantial charitable externalities are present. Alternatively, social insurance programs may be justified on the basis of one or more systems of social justice.

Summary

1 Pareto efficiency defines a situation where it is no longer possible to make mutually beneficial changes. It is Pareto efficient to exhaust all avenues for gains that benefit someone and harm no one.
2 Under theoretical conditions of perfect competition, the competitive market is Pareto efficient. In the Edgeworth box for exchange, regardless of the initial endowment position, a competitive free exchange will lead to a Pareto efficient point, a point on the contract curve. This is the First Fundamental Theorem of Welfare Economics.
3 The Second Fundamental Theorem of Welfare Economics also is illustrated by the Edgeworth box for exchange. The theorem holds that we can achieve any Pareto efficient outcome, in principle, by a competitive market, given a suitable initial distribution of resources.
4 Price discrimination is Pareto inefficient. This result includes that form of price discrimination arising when a favored segment of the population is provided reduced prices for health care to improve access.
5 The health care sector deviates from the conditions of perfect competition in many respects. These include major issues of the role of uncertainty, the role of information, and the role of externalities.
6 The Theorem of the Second Best suggests that promoting competitive features in health care markets is not welfare enhancing *per se*. Nevertheless, many competitive proposals in practice may be welfare enhancing.
7 Social programs for the subsidization or provision of health care can be theoretically rationalized on grounds of efficiency as well as equity. The usual efficiency argument posits the existence of a charitable externality in health care. The presence of externalities may, in principle, justify market interventions.
8 Need-based distributions of health care resources tend to be based on equity grounds and usually imply a rejection of market outcomes. Health care need may be understood as health care resources required to attain a given health goal chosen by society.
9 Society's optimal choice of goals may be summarized by the social welfare function, defined over all possible combinations of the individual utilities of society members. This leads to the optimal choice by selecting the point on the utility-possibility frontier that maximizes social welfare.
10 Daniels bases his concept of health care need on the fair equality of opportunity. He argues that health care needs may be identified separately from other social decisions.
11 Several need-based distributions can be criticized on economic grounds. These grounds argue that health care needs should not be (1) the technical maximum, (2) independent of cost, (3) chosen solely on technical medical criteria, or (4) monotechnic.
12 A philosophical theory of justice is needed to provide an ethical grounding for a proposed distribution. There is no present consensus on such a theory of justice.

Discussion Questions

1 At point 0_A in Figure 18.1, Belinda has all of both goods. Is this point Pareto efficient? Is it equitable? Discuss.
2 If society could clearly choose an equitable point reflecting a distribution of the two goods, is this point inevitably going to lie on the contract curve?

3 Choose an example of a health care market and identify ways in which it differs from the perfectly competitive model. Do you think that these deviations from competition could each be repaired by appropriate policies? Discuss.
4 We describe several economic criticisms of need-based distributions. Do any of these criticisms apply to Daniels's conception of health care need?
5 Speculate on how each of the three described theories of social justice would view government programs designed to provide infant and child care to the poor using tax dollars.
6 Under utilitarianism, one maximizes the total utility of society. What does this imply about the marginal utility for each person? What does it imply about the total utility for each person?
7 Suppose society determined that it must provide a minimal sustained level of health to everyone. What would this imply regarding society's expenditures on health?
8 Insufficient health care for some often is seen as a problem of insufficient income to purchase health care. Discuss two alternatives to social programs that provide health care.

Exercises

1 Prove that point *B* in Figure 18.1 is not Pareto efficient.
2 Draw an Edgeworth box like the one in Figure 18.2 but with only these details inside: the point *V* and budget line *AB* through *V*. Using indifference curves, depict the utility-maximizing choice for Abner. Now let the budget line rotate to *CD*, drawn to reflect a higher relative price of medical care. Again identify the utility-maximizing choice for Abner.
3 Is it possible to find a point on the contract curve that is not a competitive equilibrium?
4 Let Q_{opt} in Figure 18.4 represent the optimal level of health care in society under the external benefits rationale for social health care programs. Would members of society necessarily view Q_{opt} as the equitable amount of health care?
5 If all taxpaying members of society became "hard-hearted," feeling no external benefit in the health care provided to others, then what would be the optimal health care output under the external benefit rationale in Figure 18.4?
6 In Figure 18.7, which depicts the utility-possibility frontier, would society ever choose an inefficient point (inside *UU*) as the optimal point?
7 Suppose Fred has an income of $5,000 per year, and Harry has an income of $105,000 per year. If we tax $50,000 from Harry to give to Fred, will this represent a Pareto improvement for society? Why or why not? Would this improve society under some other criteria?

Notes

1 For an excellent alternative exposition of welfare economic issues, see Williams and Cookson (2000).
2 The exposition of need focuses on the consequences. Hurley (2000) describes this interpretation as strongly "consequentialist" and further describes alternative views.
3 The issue can be pursued further with Daniels's book (1985) and articles (1981, 1982).
4 Aaberge and colleagues (2010) analyze distributional measures in the presence of non-cash values.

Chapter 19

Government Intervention in Health Care Markets

In this chapter

- Economic Rationale for Government Intervention
- Forms of Government Intervention
- Government Involvement in Health Care Markets
- Health Sector Regulation and the Prospective Payment System
- The Theory of Yardstick Competition and DRGs
- Government Failure
- Conclusions

Previous chapters have revealed the scope of government involvement in health care delivery. We have seen that (1) government spending accounts for a substantial portion of all health care spending, (2) governments are deeply involved in producing as well as financing health care services, and (3) governments regulate the health care industries. In all advanced countries governments are at the center of most contemporary health care policy issues. In the United States, initiatives to control costs and increase access to care under the ACA have dominated the policy debate since the Act's passage in 2010.

The current chapter provides a framework for assessing the role of governments in health care markets. We begin with the conventional approach favored by economists, which emphasizes market failure as the rationale for government intervention. We follow with a historical review, including examples of federal, state, and local involvement in the health economy. We then focus on regulation of the hospital sector, one of the major cost drivers in all advanced economies. Finally, so that we can evaluate government programs with a more critical eye, we finish with a discussion on government failure to identify impediments in developing effective programs. Chapter 20 will provide detailed coverage of the main social insurance programs in the United States. Chapter 21 will examine health care systems in other countries, while Chapter 22 will focus on the ACA and other recent reform efforts in the United States.

Economic Rationale for Government Intervention

As we have seen, efficiency is one common standard for evaluating the desirability of economic allocations. Inefficient allocations are associated with various distortions that lead to market failure. The previous chapter introduced the distortion resulting from beneficial externalities as one justification for social programs in health care. We have also described licensure of physicians in Chapter 16 as response to information failure. Here we review the major contributors to market failure.

Monopoly Power

Monopoly power provides the classic example of market failure. A profit-maximizing monopolist produces to the level at which marginal revenue equals marginal cost. Because the marginal revenue lies below the demand curve, the monopolist will reduce production below competitive levels, and the price charged by the monopolist will exceed the marginal cost of production. The reduced production and the price–marginal cost gap together create the welfare loss.

Figure 19.1 summarizes these standard results. With a perfectly elastic competitive supply in the long run, which also represents the industry's average costs (AC) and marginal costs (MC), the competitive price and quantity are P_C and Q_C. If this industry is monopolized and no changes occur in demand or costs, the profit-maximizing output is given by Q_M, where $MR = MC$. The higher monopoly price is P_M, and triangle ABC represents the welfare loss.

However, monopoly power need not be associated solely with pure monopoly. The monopoly model is applied commonly to markets in which one or a small number of sellers are dominant. Several health care markets seem to hold a potential for the exercise of monopoly power. Examples include hospital services in markets with few hospitals, pharmaceutical products protected by patents, and some health insurance markets, often dominated by Blue Cross and Blue Shield associations.

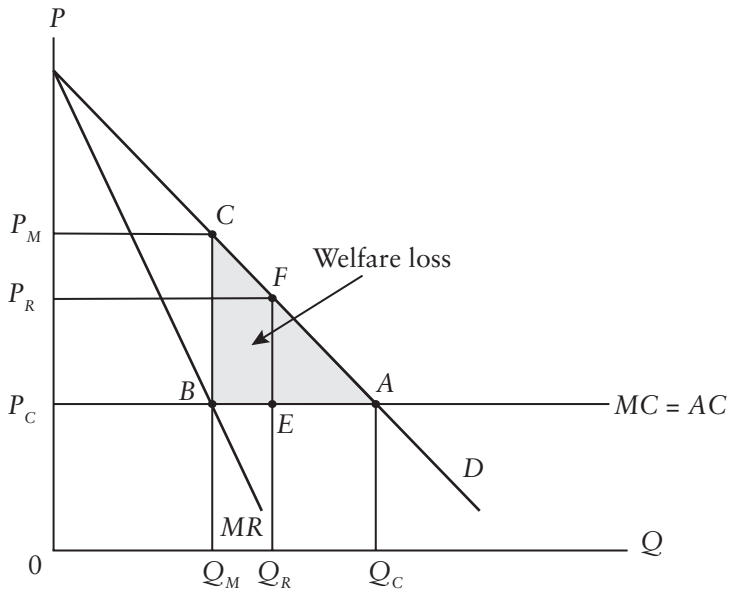


Figure 19.1 Welfare Loss of Monopoly

The potential for monopoly power exists even in markets characterized by a large number of sellers, as in the markets for doctor and dental services. Licensure laws and other forms of regulation restrict entry into some professions. Furthermore, professional associations may be able to reduce price competition by setting minimum fee schedules or by inhibiting the flow of information to buyers.

Several issues arise regarding monopoly power. First, some barriers to entry result from government intervention itself. These include licensure and patent laws. Licensure intends to ensure minimal standards of quality; patent laws seek to promote innovative activity.

Second, monopoly power may be inevitable in some situations and does not necessarily lead to economic profits. In a small market, for example, demand may be sufficient for only one hospital to survive while it just covers costs. If enforcing competition diminishes demand, even the one existing hospital might not be able to survive unless it either receives subsidies or donations, or cuts its costs.

Third, the proposed cure to monopoly inefficiencies may be worse than the problems posed by the existence and exercise of monopoly power. Some have argued that direct intervention through public provision or price controls could worsen the situation because of government failure. These critics suggest that countervailing forces and other constraints on the full exercise of monopoly power will tend to arise in private markets, especially where antitrust laws are enforced vigorously.

In the simplest case and in the absence of government failure, price controls can theoretically reduce the welfare loss caused by monopoly. In Figure 19.1, suppose a price ceiling of P_R is established. The monopolist's marginal revenue is constant, equal to the price, P_R, up to an output of Q_R. Because marginal revenue exceeds marginal cost up until output Q_R, the monopolist will produce at least to Q_R. Beyond Q_R, the marginal revenue comes from the usual *MR* curve because the monopolist would have to lower prices to attract more buyers.

Marginal revenue then will be below marginal cost and eventually will become negative. The profit-maximizing output is Q_R, and the welfare loss now falls from area *ABC* to area *AEF*. While it appears in this case that price regulation can be an effective instrument in reducing prices, monopoly profits, and welfare losses, consider that a typical hospital or physician provider may produce many different services. Consider also that demand and technology change constantly and that it may be difficult to monitor quality. As we will discuss later, price regulation under such circumstances becomes far more difficult.

Public Goods

A public good should not be confused with the public provision of a good. The postal service and local garbage collection are examples of public provision of private goods. Government may provide such goods because of natural monopoly, or a desire to subsidize certain users (e.g., rural postal customers who might not otherwise be served by a profit-maximizing postal service). In contrast, a pure public good is one for which consumption is nonrival (i.e., consumption by one individual does not reduce someone else's consumption) and nonexcludable (i.e., a consumer cannot be excluded from consuming the good either by having to pay or through some other mechanism). Economists often use national defense as an example of a pure public good. Other examples of goods having some degree of "publicness" include highways, bridges, and parks.

Market failure arises because an inefficiently small quantity of pure public goods will typically be provided without government intervention. Figure 19.2 develops this proposition. Begin with a simple case involving two persons whose demands and marginal benefits for

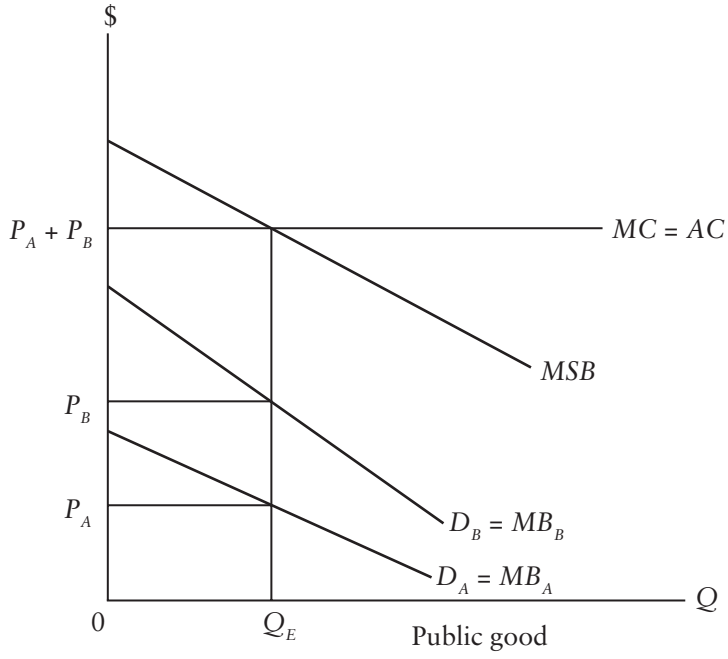


Figure 19.2 The Optimal Quantity of a Public Good

a public good are represented by D_A and D_B. Because a public good is nonrival and nonexcludable, each person must consume the identical quantity that is made available. Thus, the marginal social benefits (*MSB*) are the sum of the two individuals' marginal private benefits. Since, unlike private goods, each will get the same amount of the public good, the combined willingness of the two consumers to pay for the public good, the *MSB,* is the vertical summation of D_A and D_B.

Assuming constant costs, the efficient quantity is Q_E, where *MSB* equal the marginal costs of production, *MC*. Will the efficient quantity become available without government intervention? Knowing that their contributions are important, the two consumers may decide to cooperate by making voluntary contributions to fund at least some of the good. In fact, if each contributes according to his or her marginal private benefit at Q_E by paying P_A and P_B, respectively for every unit, enough money will be collected to cover the cost of providing the optimal quantity Q_E. It is not certain, however, whether this solution will be reached.

More realistically, public goods usually involve a large number of individuals. In principle, determining the optimal quantity, using a vertically summed marginal social benefit curve, follows the two-person example. The major difference is that these people are unlikely to cooperate to fund efficient amounts of the public good through voluntary contributions. Instead, more are likely to become free riders (i.e., make no contributions) because they cannot be excluded from consuming the good and because any voluntary contribution one makes to the provision of the good will have a negligible impact on the good's availability. The predicted undersupply of public goods in private markets has led many to conclude that governments should be responsible for making them available.

Are health care services public goods? Health services provided to one person are not consumed by others at the same time. Also, those who do not pay can be excluded from receiving care. Therefore, health care services are private goods even though they may involve public provision (e.g., through the Department of Veterans Affairs) or public financing (e.g., through Medicare and Medicaid). Thus, the public goods rationale for government provision of health care is not immediately apparent.

Despite this caveat, economic theories of public goods are highly relevant to certain health care issues. Consider the following cases:

INFORMATION Information is an economic good with a large public goods content. Especially with digitized material such as entertainment, music, newspapers, and the results of medical research, one consumer's obtaining the information does not reduce the information available to another. Although those who do not pay often can be excluded from receiving information, e.g., through copyright and patent protection, the marginal cost of providing information to another individual is relatively small. Because of this, one can argue that information will be underproduced in private markets and that government intervention is needed to increase its availability.

The government may take on two distinct roles. The first is to help disseminate existing knowledge to the public, either through direct provision or through subsidizing private sector activities. Second, governments may expand the stock of knowledge by taking an active role in scientific research, again by direct provision or through subsidizing private sector research. The federal government has pursued both strategies.

REDISTRIBUTION Voluntary giving also has a public goods dimension. Donations help raise lower-income persons' standards of consumption, including their consumption of health care. By letting others donate and knowing that one's contribution will have little impact on total contributions, individuals often choose not to pay, becoming free riders. To help offset this effect, we may need mandatory programs to correct the undersupply of voluntary giving.

However, considerable disagreement can arise over the form of the redistribution. Should redistribution occur through cash transfers to let recipients spend their additional incomes as they see fit? Or should some programs take the form of in-kind transfers (e.g., Medicaid) where recipients must use the transfers to purchase health care services? The many issues relating to equity and mechanisms for redistributing income are discussed in Chapter 18 and Chapter 20.

Externalities

In contrast to pure or nearly pure public goods, externalities refer to those goods that have third-party effects. Recall that externalities arise when a third party is affected by another party's consumption or production of a good. If a neighbor's loud music or smoke from burning leaves bothers you, it constitutes a negative externality.

Moreover, to be considered an externality, the effects must be transmitted outside the price system. An increase in demand for lower-cholesterol meats that raises their prices, adversely affecting consumers of these products, is not an externality. The higher prices ration the supply of low-cholesterol products to those who value them the most.

The major problem is that the prices of the goods and services may not fully reflect many negative or positive externalities. Thus, even when competitive forces drive prices to the marginal private cost of production, social efficiency requires that marginal social benefits equal the marginal social costs. Marginal social benefits sum the marginal private benefits and any marginal external benefits that might exist, while marginal social costs similarly sum the marginal private costs and any marginal external costs.

When a negative externality, such as pollution, creates a marginal external cost, a competitive market tends to overproduce the polluting good relative to the socially efficient quantity. Consider the case of junk food. If consumers do not bear the full cost of consuming it because the additional cost of treating the adverse health effects are passed on to others, the market price of junk food will not reflect the external cost and consumption of junk food will be too great. Conversely, competitive markets tend to undersupply goods that create beneficial (positive) externalities. Marginal social benefits exceed the price at the competitive output. Positive externalities can be important in health care, as when a charitably minded person derives satisfaction from knowing that the sick, poor, or uninsured consume more health care. More tangible externalities occur when others are inoculated against communicable diseases.

We should not confuse positive externalities with health benefits that are largely private. Subsidies that lead to improved health are often supported on the grounds that recipients will benefit society by being more productive. However, the gains from an individual's increased productivity are largely private (the individual earns higher wages and/or the employing producer earns more money, both market outcomes). Thus, the effects of the consumption of health care on productivity are not externalities.

Other Rationales for Government Intervention

Several other arguments favor government intervention. An important responsibility of the federal government is to stabilize the economy through macroeconomic policies. Although macroeconomics does not usually focus on specific sectors of the economy, changes in monetary, fiscal, and debt policy can have major effects on federal and state health care programs, as well as on private health care spending through changes in taxes and interest rates.

Another distinct category involves government's role in promoting the consumption of *merit goods*. Merit goods are commodities thought to be good for someone regardless of

the person's own preferences. Supporters of the arts, compulsory education, and mandated consumption of other goods argue that individuals do not always know what is in their best interests. Undoubtedly public policy with respect to public health interventions, such as vehicle seat belts and alcohol, tobacco, and drug use, has reflected the merit goods idea.

A final role for government involves incomplete markets which occur when private markets fail to meet an existing demand. Certain insurance markets, such as those for patients with cancer, HIV/AIDS, or other pre-existing conditions who seek new insurance, may represent incomplete markets in the sense that patients may be unable to buy insurance at any price. Government may fill these gaps by providing insurance or requiring insurers to do so.

We must determine, however, whether some of these markets are truly incomplete. Is there sufficient demand by those willing to pay actuarially fair rates so that a market would emerge? Because premiums would, on average, match insurance payments, they would be very high; this happened in the early years of the AIDS epidemic of the 1980s. Are such patients seeking subsidies by having legislation guaranteeing access to insurance at lower than actuarially fair rates?

Forms of Government Intervention

Governments can adopt a variety of policies and instruments to influence the allocation of resources or the distribution of income. The principal categories relevant to health care are selective commodity taxes and subsidies, public provision of health care, transfer programs, and regulation.

Commodity Taxes and Subsidies

We already have established that a competitive market is inefficient when beneficial (positive) externalities result from the consumption of a commodity. Using inoculations against infectious disease as an example, we extend our previous work to show how taxes and subsidies can, in principle, correct for the externality.

Let demand, *D*, and supply, *S*, in Figure 19.3 reflect the marginal private benefits (*MPB*) and marginal private costs (*MPC*) of inoculations. *MPB* equals *MPC* at the equilibrium quantity, Q_1. However, the competitive allocation is inefficient when those inoculated confer a marginal external benefit (*MEB*) on others. Too little is produced at Q_1 because marginal social benefit, the vertical sum of the *MPB* and *MEB*, exceeds marginal social costs (here equal to *MPC* because there are no external costs). Pareto efficiency and the elimination of the welfare loss shown by triangle *ABC* require output Q_2 where *MSB* equals *MPC*.

The imposition of a commodity tax or subsidy provides a method of correcting for the externality. Assume for simplicity that the marginal external benefit is constant at $5 per inoculation. To correct for the externality, Congress can provide producers with a subsidy of $5. The supply curve facing consumers will shift down by the amount of the subsidy to *S** because producers will need to receive $5 less than before from consumers to produce the quantities shown along the original supply, *S*.[1] With the new supply, the equilibrium price paid by patients decreases to P_2, and the equilibrium quantity rises to the optimum Q_2. Governments pay producers $\$5Q_2$, or rectangle P_1BDP_2.

This simple example illustrates several important features of a commodity subsidy. First, the price paid by consumers will fall by the full amount of the subsidy only when the supply is perfectly elastic, as in Figure 19.3, or when the demand is perfectly inelastic. In other cases,

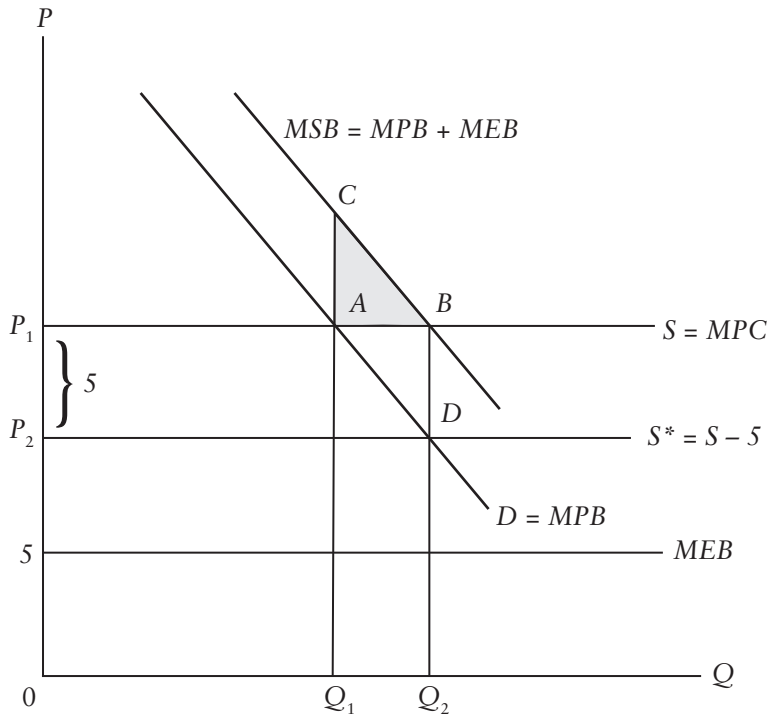


Figure 19.3 Using Commodity Subsidies to Correct for Positive Externalities

as with a positively sloped supply, the price paid by consumers will not fall by $5. Including the subsidy, producers will receive a higher price than before. Both consumers and producers share the benefits of the subsidy even though the government pays it to producers.

As the elasticity of demand increases, or elasticity of supply decreases, the equilibrium price falls by smaller amounts, and the equilibrium quantity increases by larger amounts. In the long run, the competitive market supply will be highly elastic (perfectly elastic in a constant-cost industry). Thus, the benefits of a subsidy will accrue mainly to consumers in competitive markets.

Second, to achieve efficiency, decision makers need accurate estimates of both the monetary value of the externality and elasticities of demand and supply. The costs of administering taxes and the difficulties of accurate estimation are strong arguments against the use of a corrective subsidy unless the positive externality is substantial.

Finally, we note that in theory policymakers can achieve the results described in Figure 19.3 by taxing those who are not inoculated rather than subsidizing those who are (see Exercise 4). By imposing a cost on those who refuse inoculations, the opportunity cost of an inoculation is its price minus the tax. The lower effective price increases quantity. The administrative complexity of monitoring whether people are inoculated argues against this approach.

When consumption of a good, such as smoking, leads to harmful (negative) externalities, the marginal external cost must be added to the marginal private cost to determine the

efficient solution. The competitive output will be greater than the optimum. To reduce consumption, price must be raised. A price increase can be achieved either through a corrective tax to shift the supply curve up (Box 19.1) or through a subsidy to nonsmokers that raises the effective price of smoking.

Public Provision

Roads, education, water, and police and fire protection are just a few examples of the many goods and services provided by governments. Most of these goods are not pure public goods. Although national defense often serves as an example of a good that is both nonrival (my use does not prevent your use) and nonexcludable (I cannot keep you from enjoying the good even if you don't pay for it) we cannot say the same about water, education, and most health care services. Inoculations are rival and excludable even though they may generate substantial positive externalities.

Public provision of health care is a complex process requiring a decision for each of the three basic economic questions (what? how? and for whom?) faced by every society. The "what" question relates to the types of health care to be provided (e.g., limited services, such as inoculations or comprehensive health care) as well as their quantity and quality. Whether governments themselves produce the services and how they do it, or whether they contract with the private sector, is a part of the "how" question. The "who" question deals with the financing and distribution of the services. Will the program be created as a universal entitlement, as an entitlement for some groups, or as one with other eligibility requirements? Furthermore, should the program be "free" at the point of service and funded mainly by tax revenues, or should the beneficiaries be charged user fees? The funding mechanism could have large impacts on the resources allocated to health and on the possible redistributions as from rich to poor and young to old.

BOX 19.1

Is There a Case for a Sugar-Sweetened Soda or "Junk Food" Tax?

Chapter 7 described some of the health care consequences of obesity. To the extent that individual insurance premiums do not fully reflect the additional costs of treating obesity-related diseases, and recognizing the burden that obesity places on publicly funded programs, the obese population unquestionably imposes negative externalities on others. For obvious reasons, legislating a corrective tax on individuals considered obese, i.e., Body Mass Index (BMI) > 30, is not generally practical or realistic. Nevertheless, a growing number of employers have adopted wellness programs that penalize employees who fail to take steps to manage their weight, or reward those that do so. For example, in 2008, Alabama required its state workers to undergo free screenings. Those who refused the screenings or who refused to take action for problems such as high blood pressure, high cholesterol, or obesity faced higher insurance premiums beginning in 2011. According to the Kaiser Family Foundation, about 50 percent of all firms, and 80 percent of those with 200 or more employees, offering health benefits in 2015 had wellness programs (Pollitz and Rae, 2016).

In addition to wellness programs, health care experts have directed considerable attention to "junk food." Junk food is generally associated with calorie-dense food that has high levels of sugar, glycemic starch, and saturated fat. The experts often associate such products with obesity, especially childhood obesity (e.g., Harris et al., 2009), but we caution that a causal relationship between junk food and obesity is more difficult to establish (Collins and Baker, 2009).

Although economists have examined the possibility that excess caloric consumption at restaurants is offset by reduced caloric intake at other times (Anderson and Matsa, 2011), their research has focused on demand elasticities for various junk foods and the effects of taxes on consumption. Powell and Chaloupka (2009) review the literature on food price policy and taxes in the United States. Food is subsidized for those who qualify for SNAP (formerly known as the Food Stamp Program) and several other programs such as the National School Lunch and Breakfast Program. For the most part, there are no restrictions on the types of food that are allowed under these programs. As for taxes, many states impose sales taxes on at least one category from among soft drinks, candy, and snacks (Kim and Kawachi, 2006). However, with relatively price-inelastic demands for these products, even tax rates as high as 20 percent would have little impact on annual consumption.

From a review of nine scholarly articles that met certain selection criteria, Powell and Chaloupka reach a sobering conclusion regarding the effects of food and restaurant prices on BMI and weight. The relatively small taxes on junk food which we have experienced have not produced significant changes in weight outcomes. More substantial price increases might lead to some measurable effects, especially for children and adolescents.

In the United States, numerous jurisdictions have attempted to levy additional sales taxes or impose other restrictions on junk food with the most visible effort led by former New York City mayor, Michael Bloomberg. In 2012, the city approved a regulation barring restaurants and many other types of businesses from selling sugar-sweetened drinks in cups larger than 16 ounces. Although the New York Supreme Court subsequently ruled against the ban, in 2015 Berkeley, California became the first city to impose a tax on sugar-sweetened sodas at the rate of $0.01 per ounce. The impact of the tax will depend on the extent to which it is shifted to consumers and the demand elasticity for such beverages.

On a national scale, Mexico introduced a one peso per liter tax, or roughly 9 percent of average retail price, on sugar-sweetened beverages in January 2014. Grogger's (2015) preliminary work shows that prices of regular sodas increased by 12–14 percent relative to other beverages right after the tax. With a unitary price elasticity of demand in Mexico (Colchero et al., 2015), it appears that consumption of sugar-sweetened beverages there has been significantly reduced through tax policy.

Other countries have adopted even more aggressive legislation. In 2011, Denmark introduced a tax of about $1.30 per pound of saturated fat in a product. After strong consumer resistance, including significant numbers who traveled to neighboring countries to purchase dairy products, the Danish government repealed the tax in 2012. Hungary pursued a far more ambitious strategy. To help deal with a life expectancy that is about five years below the European Union's 77-year average, in 2011, Hungary imposed a substantial tax on foods with high contents of fat, sugar, and salt. The *New York Times* reported that many Hungarians see the tax largely as a revenue-raising measure and that food manufacturers are in a constant cat and mouse game with the government as they try to reformulate products in ways to escape the tax.

While junk food taxes may raise significant revenues, analysts have found little evidence that they make any meaningful dent in obesity levels. As a result, there is interest in other interventions to replace or supplement tax policy. These include

"traffic-light" systems that rely on food labeling with visual cues on nutritional content. The United Kingdom adopted a voluntary red-amber-green color program to reflect the nutrient levels of fat, sugar, saturates, and salt in processed foods. Using a choice experiment survey to gather data on willingness-to-pay for reductions in these nutrients, Balcombe et al. (2010) found traffic-light labeling to be promising. Econometric analysis of the survey data indicated a high willingness-to-pay for a shift from red to green for all four nutrients, with much lower willingness-to-pay for shifts from amber to green, i.e., consumers are especially concerned about high levels of these nutrients.

Regulation and taxes on junk food have, not surprisingly, encountered industry opposition and legal action. But, at least in the United States, there is also substantial public opposition to taxation of sugar-sweetened beverages and candy (Min, 2013) or placing restrictions on portion sizes served in restaurants. Many consider these actions as too "paternalistic" and infringing on the freedom of choice for both producers and consumers. Mississippi, considered the most obese state in the nation (as measured by BMI), went even further in 2013 when its governor approved an "anti-Bloomberg" bill that would prohibit its "cities and counties from implementing bans on oversized food or beverage portions or mandating restaurants list calorie counts."

Sources: Suzanne Daly, "Hungary Tries a Dash of Taxes to Promote Healthier Eating Habit," *New York Times*, March 2, 2013: www.nytimes.com/2013/03/03/world/europe/hungary-experience: Accessed March 9, 2014, and Helena B. Evich, "Mexico Soda Tax to Re-ignite U.S. Debate," *Politico Pro*, January 14, 2014: plitico.com/story/2014/01/mexico-soda-tax: Accessed March 9, 2014; ICTMN Staff, "Law Bans Restricting Portion Size in Mississippi, Nation's Fattest State," *Huffington Post*, March 19, 2013: Accessed February 9, 2016.

Transfer Programs

Cash transfer programs usually are intended to meet society's equity concerns by redistributing income, with recipients free to spend their incomes in any way they want. Social Security for the elderly and some disabled is the principal example, but income supports also are provided for the poor. Cash transfers to the poor include Temporary Assistance to Needy Families (TANF), funded by the federal government and administered by the states.

In-kind transfers (benefits other than cash) also redistribute income, but their main purpose is to increase a recipient's consumption of specific goods or services. Important in-kind transfers include Medicare for the elderly, food through the SNAP program, housing, and Medicaid for those who pass means tests. Medicare and Medicaid are described in detail in Chapter 20.

Regulation

Governments influence the allocation of resources by establishing rules and regulations. In the extreme, governments can prohibit certain goods or activities entirely, such as the production and consumption of illicit drugs. More commonly, governments regulate the form or terms under which goods are produced or consumed. Regulation may appeal to legislators because it appears to tackle problems without incurring substantial government spending in the process. For example, regulating managed care to prevent "drive-through" deliveries appears

to respond to a public concern at little direct cost to government. However, as described in Chapter 12, the legislation imposes significant costs on others.

Regulation in health care markets can take many forms: licensure laws; mandates; and regulation of price, quality, and quantity of services. The following overview of the scope of government involvement in the health care sector contains several examples of regulation, with hospitals as a focal point for regulatory measures. It will be followed by sections describing various forms of hospital regulation, especially the Prospective Payment System (PPS), we will finish with a discussion of models of government failure in regulating.

Government Involvement in Health Care Markets

Government intervention in the health care markets comes mainly through three activities: provision of goods and services, redistribution, and regulation. Through public or VA hospitals, and other programs, they provide substantial amounts of health care, though this activity is overshadowed by social insurance programs for the elderly and many poor. The provision of health care and of health insurance is also the major means used to redistribute income to lower-income groups from higher-income groups.

Less obvious to many is government's role as a regulator. At the federal level, the Securities and Exchange Commission (SEC), the Environmental Protection Agency (EPA), and the Occupational Safety and Health Administration (OSHA) are regulatory agencies that affect nearly every business and working individual. In addition, states and localities impose various requirements such as those governing building and safety codes.

However, when economists and others speak of regulating or deregulating the health care industries, they are not referring to the kinds of social and commercial controls cited earlier but rather to regulations such as HIPAA (Box 19.2) targeted specifically at the health care industries. Government involvement in the health economy takes on many forms, some of which are developed elsewhere in this text. Here we provide examples to highlight the variety and scope of government intervention in health care markets.

BOX 19.2

What Is HIPAA?

The Health Insurance Portability and Accountability Act (HIPAA), that passed in 1996 and required compliance by October 2003, is one of the most far-reaching health care regulatory measures ever approved. A principal goal of HIPAA is to protect workers for loss of coverage when they change jobs, especially to reduce "job lock" for those with pre-existing conditions. It also prohibits discrimination by insurers based on health status. Among other goals, including standards for electronic data interchange, the Act also was intended to guarantee the security and privacy of patient health data. Patients now "own their records" and must be able to access their records and know how their personal information will be used. Changes in 2009 introduced guidelines on how to protect health information appropriately. In 2013, new requirements dealing with privacy, security, and breach notification were introduced. Rules and guidance on securing protected health information through encryption were among the most important changes.

HIPAA requirements are complex, and the penalties for failure to comply are stiff. As a result of the complexities, a cottage industry consisting of legal experts, information specialists, and consultants has emerged to help physicians, hospitals, insurers, and other providers to achieve and maintain HIPAA compliance. The economic effects of HIPAA requirements on costs and other consequences, including restrictions on the ability of researchers to access health care data, have not yet been fully evaluated.

Support of Hospitals

As described in Chapter 14, the modern hospital did not begin to emerge until the confluence of several developments in the late nineteenth century, such as major improvements in anesthesia, antisepsis, and in the invention of X-rays. Temin (1988) characterizes hospitals prior to this period as being more like municipal almshouses funded by taxes or voluntary contributions. Hospitals "existed for the care of marginal members of society, whether old, poor, or medically or psychologically deviant" (pp. 78–79).

In retrospect, one can argue that public support for hospitals reflected a redistribution motive and a desire to deal with the negative externalities associated with living with the insane and those harboring communicable diseases, such as tuberculosis. With the improvements in physicians' abilities to diagnose and to treat patients surgically, hospitals grew rapidly in the first decades of the twentieth century. Public hospitals continued to serve the poor but also focused their attention on the growing middle classes. Ultimately, patient payments and insurance became the primary sources of funds for many of these institutions.

Federal support for private hospitals was minimal until the passage of the Hill-Burton Act in 1946. At that time, about 40 percent (over 3,000) of U.S. counties did not have a community hospital, with many existing hospitals considered substandard. The Hill-Burton Act sought to expand rural health facilities by providing for matching grants to nonprofit institutions. The program, which helped finance about one-third of all hospital projects, contributed substantially to the rise in hospital beds per capita between 1947 and 1975 when funding under the Act ended.

Department of Veterans Affairs and Department of Defense

Governments are major providers of health care. Federal, state, county, and municipal hospitals account for approximately 17 percent of total hospital beds in the United States. The largest public provider is the Department of Veterans Affairs (VA). In 2013, the VA spent $55 billion for health care that included 90 million outpatient visits and 632,000 hospital discharges. The VA's primary purpose is to provide care for service-related injuries through institutions that specialize in providing and undertaking research for such care. However, it will treat veterans for other conditions unrelated to service injuries if facilities are available and the veteran indicates an inability to afford treatment from other sources. As a result, most patients in VA hospitals are lower-income people treated for conditions not related to their military injuries. A large portion of the substance abuse treatment in the U.S., for example, comes through the VA system.

In addition to direct care provided by the VA, the Department of Defense (DoD) provides coverage to active-duty service members as well as reservists and various other categories of

military personnel, their families, and survivors. Coverage comes through its TRICARE program which provides several options including HMO and fee-for-service plans. TRICARE spent over $52 billion in 2012 to provide coverage for almost 10 million service members, retirees, and their families. Health spending, as a share of total DoD spending, increased from 6 percent to nearly 10 percent in 2012. While much of this increase is due to growth in the number of beneficiaries, Congress has been concerned about the cost burden facing the DoD. In 2016, President Obama called for limiting TRICARE to two plan alternatives and requiring all retired service members to pay an annual enrollment fee.

Food and Drug Administration

As noted in Chapter 17, the U.S. has passed several acts of drug legislation beginning with the Food and Drug Act of 1906. Legislation in 1962 required increased testing and evidence of efficacy and increased the control of the Food and Drug Administration (FDA) over the introduction of new products. Among many provisions found in the most recent amendment (in 2007), the FDA's authority and resources to conduct reviews were considerably expanded.

The pharmaceutical industry believes that regulatory requirements lengthen the period to FDA approval to about 10 to 14 years and raise the cost of introducing new drug products, estimated by DiMasi and colleagues (2003) to be as high as $800 million (in 2000 dollars) to the point of marketing. The economic issues of drug regulation pit the relative gains in drug safety and efficacy against the discouragement to innovation and the delays in availability attributable to the approval process.

Mandated Health Insurance Benefits

The insurance industry is mainly regulated at the state level. The states traditionally seek to ensure financial solvency of insurance companies, but in both auto and health insurance the states have expanded their roles considerably. One relatively recent phenomenon in health insurance is the proliferation of mandated benefits. Mandates can require coverage for specific health care services, such as breast reconstruction; provider types, such as physical therapists; or even who is eligible for coverage, such as domestic partners. Such laws have exploded from 48 in 1970 to almost 1,100 in 2016. They often pass after lobbying by previously excluded special interest groups, or after highly publicized cases involving persons who find they are not covered for heart transplants or certain expensive experimental therapies.

With the passage of the ACA and its ten categories of essential health benefits, much of the action on mandates has shifted from the state to the federal level. Chapter 11 describes those essential benefits in detail and analyzes their wage and employment effects. The ACA also introduced the popular mandate requiring all employer plans and those purchased in the individual market to cover dependent children (whether single or married) through to age 26.

Tax Policy

Federal and state governments provide substantial tax subsidies for the consumption of health care and purchases of insurance. In particular, employer contributions to group health plans are not included in an individual's taxable income (escaping federal, state, and Social Security

taxes). Individual payments for health insurance as well as many health care expenses can be itemized if they exceed 7.5 percent of adjusted gross income in 2015 for those over 65 and 10 percent for those over 65.

Some have argued that the substantial reductions in the after-tax prices of insurance and health care represent a major reason for the growth of insurance and consequent growth of health care spending that began in the 1960s and 1970s. Thus, a proposed cost-containment strategy is to reduce such subsidies. Without discussing the merits of this argument here, we point out that there is little justification for the extensive tax subsidies. Ironically, the tax advantages accrue most to those in the highest marginal tax rates, meaning that the subsidies do not promote equity.

Public Health

We have thus far focused largely on the study of personal health and health care decisions. Public health, in contrast, deals with communicable diseases, epidemics, environmental health issues, and other activities with significant third-party health effects, such as smoking and the use of drugs and alcohol.

Until the latter part of the nineteenth century, public health was a state and local responsibility. However, in 1878, the U.S. government created the Marine Hospital Service, which became the U.S. Public Health Service in 1912. Total federal spending on public health activities was $79 billion in 2014 with another $11 billion in state and local spending. Odin Anderson (1990) provides an excellent history of public health in the United States; Mays and Smith (2011) provide current estimates on the significant reductions in preventable causes of death associated with increased public health investments.

Other Government Programs

Numerous other channels for government involvement exist. Some are obvious—such as support for medical education and medical schools, which will influence the supply of providers. Similarly, support for health care research undertaken directly by a government agency, such as the National Cancer Institute, or undertaken by other public and private organizations can have a substantial impact on the spread of technology and the direction of the health care system. For example, critics of the U.S. system claim that it has placed excessive emphasis on high-tech medicine at the expense of preventive medicine and that this bias stems in part from the kinds of research projects supported.

Other forms of intervention are rather less obvious. Changes in immigration policy can affect the supply of health care personnel at all skill levels. Also easily overlooked is the federal government's role as the nation's largest employer in providing fringe benefits to its employees. The Federal Employees Health Benefits Program (FEHBP) offers numerous insurance options to employees and their dependents. Because of its size and its ability to experiment with alternatives, the FEHBP has the potential to influence and serve as a model for the private sector.

To summarize, in the United States governments at all levels have intervened heavily in the health care sector. The intervention has taken the forms of direct provision of care, financing of health insurance (especially of Medicare and Medicaid), subsidizing of medical education and the construction of health facilities, subsidizing of purchases of health insurance through the tax structure, and the regulation of health care industries. This intervention has major effects on the total resources devoted to health care as well as on the distribution of resources within the health care industries.

Table 19.1 National Health Expenditures, by Sponsor, Selected Years

Type of Sponsor	*2008*	*2011*	*2014*
National Health Expenditures ($ billions)	2,403	2,697	3,031
Businesses, Household, and Other Private Revenues	1,412	1,506	1,673
Private Businesses	514	547	606
Household	725	778	844
Other Private Revenues	173	182	222
Governments	991	1,191	1,359
Federal	581	731	844
State and local	410	460	515
Government's share of NHE (%)	41.2	44.2	44.8

Source: Martin et al. (2016, Exhibit 5).

Table 19.1 shows national health expenditures by sponsor for 2014. The Centers for Medicare & Medicaid Services (CMS) defines *sponsor* as the "entity that is ultimately responsible for financing" the health care spending. Governments financed 45 percent of the $3.03 trillion in national health expenditures, with the federal government's share at 62 percent of the $1.36 trillion government component. In the brief period since the passage of the ACA, the government's share has been relatively steady but analysts at CMS project that additional obligations will increase the share to 47 percent by 2024 (Keehan et al., 2015). These obligations include expanded Medicaid eligibility, growth of the Medicare population, and federal tax and premium subsidies for insurance coverage for those between 100 and 400 percent of the federal poverty level.

Expenditures for the largest government programs, Medicare and Medicaid, amounted to $619 billion and $496 billion, respectively, in 2014. The federal government was responsible for 61.5 percent of Medicaid spending, while the states picked up the remaining 38.5 percent. Other important federal programs include public health, research, the Departments of Defense and Veterans Affairs, and the Children's Health Insurance Program. Expenditures for all these programs totaled about $174 billion in 2014.

Health Sector Regulation and the Prospective Payment System

We have established that most health care providers (hospitals, clinics, physicians) face portions of downward demand curves because patients do not or cannot shop among providers, and do not immediately abandon the provider who raises its price. As such, the desirable outcomes of competitive markets become much more difficult to achieve. Using competitive markets as a comparison group, consider three major categories of policies:

1 Recognize providers' monopoly powers and try to control them. These policies include some traditional features, such as utilization review, capital controls such as Certificate-of-Need (CON) regulations, or rate controls.

2 Prevent the accumulation of monopoly power. Most antitrust policies fall into this category.
3 Make monopolistic firms act like competitors. These policies include prospective payment plans for physicians and hospitals.

Traditional health economics texts (and earlier editions of this one) gave particular weight to items 1 and 2. In terms of controlling monopoly powers, policymakers (most often at the state level) sought to regulate provider behavior by attempting to control utilization, control rates, or control the construction of (expensive) new facilities. Most evaluations of these policies showed modest and often short-lived impacts.

Antitrust policies have been directed most often at hospital mergers. Prospective merger partners argue in terms of inadequate demand or inefficient scale. Opponents argue that merging two hospitals, for example, replaces competition with potential monopoly power. Traditional antitrust criteria have included industry descriptors, four- or eight-firm concentration ratios, or Herfindahl-Hirschman indices (see Chapter 17), rather than perceived provider behavior. Over the first 15 years of the twenty-first century, particularly in the face of rapidly changing market structures due to improved transportation, and the decreased costs of searching among potential competitors, antitrust enforcement, particularly with respect to hospitals, has decreased.

We will concentrate on item 3, making firms compete, for two reasons. First, hospital prospective payment schemes through programs like Medicare have dominated the regulatory activity in the United States and in other advanced countries. Second, the passage of the Affordable Care Act has changed the regulatory environment in which insurers and providers operate. We will discuss those impacts in Chapter 22.

The 1965 enactment of Medicare and Medicaid launched the federal government into the reimbursement of services. Program costs increased substantially over the years, and as costs grew, so did interest in cost-containment policies.

Hospital payments and physician payments comprise the two major cost centers. Hospitals and physicians often bill patients separately and receive payment separately. We will focus on hospital payment in this section. Physician payment regulation in the United States has sought to realign incentives to emphasize general practitioners and de-emphasize specialists, without major success. The previous physician payment regulation program, called the Sustainable Growth Rate, was eliminated by Congress in 2015, largely because it had to be overridden almost annually. We elaborate on this more in the section on Government Failure at the end of this chapter.

Here, we focus on the successful Medicare program that converted the financing of Medicare hospital care to a prospective payment system based on Diagnosis Related Groups (DRGs). We discuss the system in the present context because its widespread use in almost all hospital reimbursement stems from its adoption by the Medicare program. However, the PPS under DRGs has wider significance to other segments of the health sector, as well as to the regulation of industry in general.

Numerous advanced countries, including Australia, Germany, England, and France, have also implemented DRGs in various forms. Even less-advanced countries from Estonia to Mongolia are planning implementation (see Mathauer and Wittenbecher, 2012 for more details). Although each country's plans differ, we will see that the core organizing principles are similar.

Description of PPS

Contrast the current prospective payment to Medicare's previous retrospective reimbursement system. Under retrospective payment, a hospital submitted its bill to Medicare after the care was provided and the costs to the hospital were known. Retrospective payment allowed the hospitals to recover their expenses as allowed by Medicare rules whether these expenses were high or low, excessive or efficient. Retrospective reimbursement provided only modest incentives for hospitals to control costs.

Prospective payment, in contrast, sets payment rates *prior* to the period for which care is given. By setting a *fixed* reimbursement per admission, prospective payment provides economic incentives to conserve on the use of input resources. Hospitals that use more resources than covered by the flat rate lose the difference. Those with costs below that rate retain the difference.

Payment from the government is complicated. As of 2016, Medicare has 19 different payment systems. We group them as follows:

1 Inpatient acute care in short-term hospitals and psychiatric facilities.
2 Ambulatory care furnished by physicians, hospital outpatient departments, ambulatory surgical centers, and clinical laboratories.
3 Post-acute care furnished by skilled nursing facilities, home health agencies, inpatient rehabilitation facilities, and long-term-care hospitals.
4 Dialysis services furnished in outpatient centers and hospice care.
5 Ambulance services and products furnished by durable medical equipment suppliers.
6 Services furnished by private health plans under the Medicare Advantage program.
7 Services furnished by accountable care organizations (ACOs).
8 Part D voluntary drug benefits.

In 2007–2008, CMS adopted a new set of 745 Medicare Severity Long-Term Care Diagnostic Related Groups (MS-DRGs) that replaced the existing 538 DRGs with ones that better recognized illness severity. These DRGs seek to:

- Improve the accuracy of Medicare's inpatient hospital payments by using hospital costs rather than charges to set rates.
- Adjust payment to recognize better the severity of illness and the cost of treating Medicare patients by increasing payment for some services and decreasing payment for others.
- Eliminate biases that had provided incentives for physician-owned specialty hospitals to treat the healthiest and most profitable cases, leaving the sickest and least profitable patients to general acute care hospitals.
- Refine the payment system to provide hospitals with incentives to invest in service areas based on the clinical needs of their patients rather than financial incentives.

Each DRG has a flat payment weight calculated on the basis of costs incurred for that DRG nationally. For example, based on Table 19.2, DRG 001, for a heart transplant, has a larger weight and is about 44 percent more costly (25.3920/17.6399) than DRG 003, the use of an artificial lung (membrane) located outside the body (extracorporeal). It is over 147 times as costly as a normal newborn birth (DRG 795 = 0.1724).

Table 19.2 Fiscal Year 2015 DRGs by Weight—Five Highest and Five Lowest

MS-DRG	*Type*	*MS-DRG Title*	*Weights*	*Geometric Mean Length of Stay*[1]	*Arithmetic Mean Length of Stay*
Five Highest					
001	Surgical	HEART TRANSPLANT OR IMPLANT OF HEART ASSIST SYSTEM W MCC	25.3920	28.6	36.8
003	Surgical	ECMO OR TRACH W MV 96+ HRS OR PDX EXC FACE, MOUTH & NECK W MAJ O.R.	17.6399	26.3	32.2
002	Surgical	HEART TRANSPLANT OR IMPLANT OF HEART ASSIST SYSTEM W/O MCC	15.6820	15.7	18.7
927	Surgical	EXTENSIVE BURNS OR FULL THICKNESS BURNS W MV 96+ HRS W SKIN GRAFT	15.5499	22.9	28.9
215	Surgical	OTHER HEART ASSIST SYSTEM IMPLANT	15.4348	11.2	17.2
Five Lowest					
894	Medical	ALCOHOL/DRUG ABUSE OR DEPENDENCE, LEFT AMA	0.4450	2.1	3.0
298	Medical	CARDIAC ARREST, UNEXPLAINED W/O CC/MCC	0.4227	1.1	1.2
782	Medical	OTHER ANTEPARTUM DIAGNOSES W/O MEDICAL COMPLICATIONS	0.4057	1.7	2.3
780	Medical	FALSE LABOR	0.2880	1.2	1.4
795	Medical	NORMAL NEWBORN	0.1724	3.1	3.1

Notes: [1] The *geometric mean* or average is an alternative measure of central tendency, the *n*th root of the product of the observations. For example, the arithmetic mean of observations *a* and *b* is $(a + b)/2$, while the geometric average is $\sqrt{ab}$.

Also MCC refers to major complications or comorbidities; O.R. refers to operating room.

DRGs are very complex, like the production systems that they regulate, and some features of the DRG system do allow reimbursement to vary with actual treatment decisions during an admission. Moreover, a look at alternative methods available on the Internet for calculating hospital-specific DRG-adjusted payments shows how truly complicated the procedure is, with adjustments for differences in hospital labor costs, disproportionate shares of low-income clients, and "hold harmless" clauses (reflecting prior practices). Most importantly, however, the rates are flat in that they do not change for hospitals that spend more than the rate or, for that matter, less.

The Theory of Yardstick Competition and DRGs

Shleifer (1985) describes the theory of yardstick competition, a close approximation to PPS under DRGs. We can think of yardstick competition as the ideal form, while the actual Medicare payment system is a real-life approximation. Shleifer considers markets where firms have monopoly power. Most medical providers face downward-sloping demand curves so they possess some degree of monopoly power.

Recall that monopolistic firms reduce outputs and correspondingly charge higher prices than perfectly competitive firms, Moreover, many analysts believe that secure monopolistic firms, without the market discipline of competition, may lack some incentives to reduce costs. In U.S. hospital markets prior to PPS, one of the most serious cost-efficiency disincentives was the retrospective payment system already discussed. Shleifer's yardstick competition describes a regulatory scheme, much like Medicare's PPS, that restores cost-consciousness incentives.

Return to Figure 19.1 in which the hospital faces demand curve D, and where marginal and average costs equal P_C. By equating marginal revenue and marginal cost, the monopoly hospital will provide Q_M units of output and charge P_M indicated by point C, with an initial economic profit as indicated in Box $P_M P_C BC$. If instead, the hospital received reimbursement prospectively at a rate of marginal cost P_C, the hospital would produce quantity Q_C at the intersection of demand D with P_C, at point A.

Suppose the hospital guesses that it could produce at lower cost if it hires a team (a *fixed cost* because it is unrelated to output) of efficiency experts and carries out its advice (also a fixed cost). Recall that economic efficiency requires a firm to produce a quantity at which marginal cost (value in production) equals the market price (value in consumption). The problem for yardstick competition designers is to set up a payment scheme so that these firms have the incentives to spend just the right amount of money and effort on reducing production costs.

What is just the right amount of expense to incur in the effort to reduce production costs? Suppose that a hospital treating 5,000 cases per year discovers that it could reduce its marginal costs of production by $100 per case treated if it spent $150,000 on cost-reduction efforts. Would the cost-reduction effort be worth it? Consider the first line of the following schedule:

Each step will reduce costs per case by $100 and step (1) represents the first of four possible steps. With step (1), reducing per-unit costs by $100 costs a total of $150,000 in cost-efficiency efforts. The extra $1 saved for every case treated generates $500,000 in extra revenue.

This step of cost-saving is worthwhile because it costs less ($150,000) than it saves ($500,000). By similar reasoning, one more step of cost-saving also would be worthwhile, costing $270,000, but saving another $500,000. Step (3) is worthwhile, too, but that is where we would stop. Society's problem is that if competition provides incentives to cut costs, unregulated monopolists or retrospectively reimbursed firms may not have the incentives to

take these three steps. It would be good to design a payment system that would induce them to do so.

How does yardstick competition provide this inducement? Shleifer constructs an *economic game*. Games occur when firms must engage in strategies contingent on what other firms do. In many geographic markets, there are relatively few hospitals or hospital systems, so each hospital must take the decisions of other hospitals into account in determining its best strategy.[2]

	Marginal Cost of Efficiency Effort		*Marginal Surplus (Revenue) Generated*
Step (1)	\$150,000	<	\$100 × 5,000 = \$500,000
Step (2)	\$270,000	<	\$100 × 5,000
Step (3)	\$490,000	<	\$100 × 5,000
Step (4)	\$750,000	>	\$100 × 5,000

Consider Hospitals A, B, C, D, and E, in a metropolitan area. If reimbursed for its own efficiency efforts, Hospital A (and Hospitals B, C, D, and E) has the incentive to state high costs for its efforts—in the table above, maybe \$1,000,000. Shleifer's mechanism, however, sets the reimbursements $\bar{R}$ equal to the averages of the marginal costs of *all other hospitals* in the market. So, if Hospital A (and the others) "highball" their estimates, they put themselves at a disadvantage because their competitors get larger reimbursements—and *they don't*.

To get to the efficient output, the regulator must set the regulated price at firms' improved (lowered) marginal costs. If the hospital can convince regulators to assign it a high regulated price, it can increase its earnings. Once again Shleifer's mechanism, however, sets the prices equal to the averages of the marginal costs $\bar{C}$ of *all other* hospitals in the market. Yet again, if Hospital A (and the others) "highball" their marginal cost estimates, they put themselves at a disadvantage because their *competitors* are allowed to charge higher prices. So, the game sets the hospitals' greed against them. If they charge too much, others benefit. If they charge too little, they lose money.

What do the hospitals do to try to win the game? Under the yardstick mechanism, hospitals compete, and this competition leads overall to price equaling marginal cost. The competition, known as a "Nash equilibrium" (after the late Nobel Laureate John Nash), refers to a market solution in which each firm does the best that it can, given the decision of others. It is an *equilibrium* because once the choices are made, no firm has any motive to change its action. Shleifer views it as

> essential for the regulator to commit himself not to pay attention to the firms' complaints and to be prepared to let the firms go bankrupt if they choose inefficient cost levels. Unless the regulator can credibly threaten to make inefficient firms lose money . . . cost reduction cannot be enforced.
>
> (p. 327)

Suppose that Hospital A's new marginal cost C_A^* fortuitously equals $\bar{C}$, the average of its competitors. Under yardstick competition, it will lose an amount equal to the fixed costs of its efficiency efforts. A lump-sum subsidy from the regulator to the hospital will defray some or all of the fixed costs required for these efforts.

If the hospital has succeeded in lowering its marginal costs below $\bar{C}$, then it will earn a profit. If the hospital cannot lower its costs as low as $\bar{C}$, then it will likely lose money. In either case, because the hospital's actual costs *do not* enter into the price that it receives,

and because the hospital can earn a profit if it reduces costs, the hospital has a considerable incentive to do so.

Return now to DRGs as practiced. Consistent with Shleifer's formulation a hospital's actual costs do not enter into its formula for payment rates, so hospitals must become price takers in the strictest sense. If it costs the hospital more to provide the service than the DRGs allow, the hospital either loses money on the service and stops offering it, or cross-subsidizes the service from other services that it can produce at costs lower than their DRGs. Thus, the cost-cutting incentives are strong. In fact, following the introduction of DRGs, between 1985 and 2002, approximately one-third of the hospitals in nonmetropolitan areas ceased offering obstetrics services (Zhao, 2007). Although yardstick competition applies to the hospital, many of the costs reflect orders from physicians who are generally not even their employees.

How accurate are the yardstick payments generated for DRGs? Although designed to elicit cost-cutting behavior, we would expect DRGs on average to equal provider costs. Ginsburg and Grossman (2005) report that Medicare, and other payers, have found it difficult to devise payment rates that closely follow relative costs. Medicare prospective payment for inpatient care, based on DRGs, determines the relative payment for each DRG on the basis of average charges for patients in that DRG across all hospitals. To the degree that the pattern of charges systematically diverges from the pattern of unit costs, then relative payments for different DRGs will not reflect relative costs. As a result, patients in some DRGs are more profitable than others.

The advent of DRGs immediately led to shorter hospital lengths of stay, and this impact has continued. According to the Medicare Payment Advisory Commission (MEDPAC, 2016, p. 185), discharges for short inpatient stays have declined rapidly in recent years. Between 2006 and 2012, the number of one-day inpatient stays declined 23 percent per Medicare Part A beneficiary, a more rapid rate of decline than for longer stays (Table 19.3).

Table 19.3 Recent Impacts of Prospective Payment

		Inpatient discharges by length of stay		
		Percent change in number of inpatient claims per Part A beneficiary		
Length of inpatient stay (in days)	*Share of inpatient discharges, 2012 (%)*	*2006–2012 (%)*	*2006–2009 (%)*	*2010–2012 (%)*
1	13	–23	–10	–13
2	16	–6	–1	–5
3	19	–1	1	–3
4	13	–12	–4	–6
5 or more	39	–17	–8	–7
All	100	–13	–5	–7

Note: Hospitals receiving inpatient prospective payment system payments and critical access hospitals are included in this analysis.

Source: MEDPAC, 2016, Table 7–4.

From 2006 to 2009, the volume of one-day inpatient stays decreased 10 percent compared with the 13 percent decline from 2010 to 2012. Inpatient stays of other lengths also demonstrated an increased rate of decline between these two periods. Providers reduced inpatient utilization during the two periods, substituting (less expensive) outpatient care.

Government Failure

Market failure is a necessary condition for government intervention. As we have seen, economists associate market failure with monopoly power, externalities, and public goods, including the public goods characteristics of redistribution and information. Government policies, in principle, can correct misallocations resulting from market failure. To do so, governments can use specific commodity taxes and subsidies, public provision of goods and services, transfer programs, and regulation.

We have to ask whether government in practice can improve efficiency and better meet society's equity objectives. Many might argue that this question has already been answered. Governments in each of the 34 countries that belong to the Organization for Economic Cooperation and Development (OECD) are heavily involved in their health economies, to the point where nearly all have universal health care coverage. The United States had the second lowest public share of total health care spending in 2013 among the 34 countries, but the public share according to OECD data is still substantial—48.2 percent.[3]

The real issues are the extent and forms of government involvement. The difficulties of agreeing on objectives, choosing from many different policy instruments, and selecting the correct values of these instruments create many opportunities for "government failure."

The literature on public choice illustrates many of the problems in developing and implementing policy. Public choice attempts to model how decisions are made through the political process. While many models of public choice have been created, we limit our discussion to two features that are relevant to efficiency: special interest groups and bureaucratic behavior.

Who Does the Regulator Represent?

Wherever we have addressed regulatory activities such as licensure or other quality controls, we have assumed that the regulator knows the right actions to take. Most often we treat the regulator as representing an omniscient but benevolent despot who knows what is good for the economy and regulates accordingly. Yet historians of railroad regulation might argue that the railroads played major roles in influencing their regulators. Likewise, some observers believe that the American Medical Association has greatly influenced laws on licensure and legislation on Medicare, and the insurance industry had a prime role in the formulation of the Affordable Care Act. Can the theory of regulation address issues such as these?

It can. An influential school of economists often associated with the University of Chicago has argued that the regulation process, like many others that we have examined, is a "maximizing" process, in which a regulator seeks "votes" from a group of potential beneficiaries. These votes would allow the transfer of wealth, such as monopoly profits, from those regulated to those who benefit from the regulations.

In this model formulated by Peltzman (1976), the regulator seeks the "votes" of supporters by imposing a percentage "tax" on those who are regulated, with the tax receipts being transferred to the beneficiaries. A higher tax rate on those who are regulated (more stringent

regulation) may win more supporters, but it also may mobilize more opponents. The successful regulator seeks a certain transfer of wealth while gathering votes to effect this transfer efficiently.

Suppose the regulator starts with a modest tax rate. The imposition and increase of this rate yields favorable votes for two reasons:

1 The revenues increase the probability of support from the potential beneficiaries.
2 The revenues increase the value of the transfer to the beneficiaries.

The marginal benefit to the regulators is shown as the left-hand side of equation (19.1). As the tax rate rises, the marginal benefits fall for two reasons. There may be fewer supporters to attract (term [1]), and the taxation or regulation may be reducing the wealth left for the beneficiaries to tax (term [2]).

Expressing this opposition effect in dollars, the regulator's optimum is found from the usual condition where the marginal benefits from regulation equal the marginal costs, or:

$$\begin{array}{c}\text{mgl increase}\\ \text{in probability}\\ \text{of support}\\ \text{[1]}\end{array} \times \begin{array}{c}\text{mgl revenue}\\ \text{product from}\\ \text{those regulated}\\ \text{[2]}\end{array} = \begin{array}{c}\text{mgl}\\ \text{opposition from}\\ \text{increased taxes}\\ \text{[3]}\end{array} \qquad (19.1)$$

Opposition also will be heard. Those who would be regulated will oppose a transfer because it is coming from them. Their opposition will grow the higher the tax (the more stringent the regulation), that is, the more that is being taken from them.

The two terms on the left-hand side of equation (19.1) suggest that regulation will be sought if there is a large group of supporters (term [1]) or if there is a smaller, well-organized group that perceives large expected gains (term [2]). Thus, one may see regulations imposed that benefit special interest groups, that is, relatively small but well-organized groups of potential gainers.

While this analysis does not negate the economic arguments in favor of regulation, it provides sobering insights into how regulations may be implemented. Regulators may respond to similar economic incentives as do other economic men and women. The resulting regulations, although possibly addressing issues of economic efficiency, most likely will reflect the particular preferences of the beneficiaries, moderated by the opposition of those who stand to lose under the regulations.

We close with a recent example in the context of health sector regulation in the United States, the regulation of physician payments. Physician payment is a major cost in the health economy (along with hospitals and drugs). Efforts from the late 1980s sought both to control payment levels and to realign payments so as to emphasize general practice and de-emphasize specialist care.

Wynne (2015) describes the program called the Sustainable Growth Rate (SGR) which boosted physician payments when the growth rate of spending on physician services fell short of growth in the gross domestic product (GDP) and cut payments when physician spending grew more rapidly than GDP. Prices, the number of Medicare beneficiaries, and changes in law all entered the payment mechanism, essentially leaving utilization rate as the only key factor driving the SGR algorithm.

In 2002 the formula led to a proposed cut to Medicare's base payment rate for physician services of 4.8 percent. Physicians complained bitterly and threatened to refuse new patients.

The U.S. Congress that year, and almost every year thereafter, passed laws to prevent mandated decreases. Finally, in 2015, Congress eliminated the SGR algorithm entirely.

What happened? Many analysts believe that in the face of a myriad of complicated rules, doctors started doing more work to offset their stagnant wages in order to keep their income levels constant. Moreover, they were organized enough to threaten action against the SGR program and to convince Congress that they were serious (return to equation 19.1). Congress repeatedly postponed the mandated rate cuts. Since the repeal of SGR, a host of stopgap programs have been put together to regulate physician payment, including automatic increases for all doctors from 2015 through 2019. A sober assessment would indicate that physician payment regulation has been nowhere as successful as the prospective payment hospital regulation under DRGs.

Conclusions

This chapter emphasizes market failure as the economic rationale for government intervention. Monopoly power provides the classic example, but public goods and externalities are two additional categories that are relevant to health care. Government policies, in principle, can correct misallocations resulting from market failure. Governments can use specific commodity taxes and subsidies, public provision of goods and services, transfer programs, and regulation.

Through a wide variety of programs, governments at all levels have become major players in the U.S. health economy, accounting for 45 percent of national health expenditures in 2014. However, government activities also are associated with government failure. Information deficiencies, the efforts of special interest groups, and bureaucratic behavior can lead to socially undesirable programs or inefficient levels of these programs. Nonetheless, despite concerns about the effectiveness of government programs, governments will almost certainly continue to dominate the health economy through their various spending programs and heavy regulation.

At the same time, the United States has promoted competitive strategies to deal especially with the cost and access concerns. With the strong political pressure from some groups to downsize the role of government and even rescind the ACA, there will likely be continuing interest in market-driven changes to the private insurance system as well as to Medicare and Medicaid.

Principal regulatory mechanisms used in the United States have included rate regulation, utilization review, capital constraints, and antitrust law. The most effective ones relate to prospective payments to hospitals, which are theoretically sound, and have been successful in reducing hospital length of stay without adverse long-term patient consequences.

The following three chapters will further examine the role of governments in health care. Chapter 20 will concentrate on the main public insurance programs: Medicare, Medicaid, and the Children's Health Insurance Program (CHIP). Following an examination of health systems in other countries in Chapter 21, and the lessons learned from them, we will address health system reform in Chapter 22. We will focus on reforms directed at attaining universal or near-universal coverage in the United States, with special attention given to the ACA. Because many respected scholars and political leaders embrace market solutions to the challenges of access, costs, and quality, we will also evaluate competitive approaches to health reform.

Summary

1 Federal, state, and local governments were responsible for 45 percent of total health care expenditures in 2014. The share is projected to increase to 47 percent by 2022.
2 Governments' share of total health care spending in the United States is the second lowest among the 34 OECD countries.
3 The traditional rationale for government intervention is market failure. Sources of market failure include monopoly power, externalities, and public goods.
4 There is a public goods aspect to information and redistribution that can be used to justify a role for government in health care.
5 Commodity taxes and subsidies, public provision, transfer programs, and regulation are the principal policy instruments used by governments.
6 Regulation refers to the use of nonmarket means to affect the quality, price, or quantity of a good or service. The principal categories of regulation include fee controls and rate regulation, quantity and capacity controls, and quality controls.
7 Governments have participated in a wide variety of activities, including the direct provision of health care, subsidizing the production of health care, the provision of social insurance, public health, and regulation of health care products and providers.
8 Economists view health services regulation as desirable when competitive market pressures are not present. One can categorize the policies as those that:
   - Recognize the monopoly power and try to control it.
   - Try to make monopolistic firms act like competitors.
   - Attempt to prevent the accumulation of monopoly power.
9 The Medicare Prospective Payment System (PPS) based on DRGs predetermines a flat fee per case. Hospitals that exceed this rate suffer losses, while hospitals with case costs below the rate receive profits.
10 Medicare's PPS approximates yardstick competition. By setting the payment rates according to industry average marginal costs, yardstick competition induces the firms to choose the socially efficient level of cost-containment expenditure.
11 Medicare's PPS has reduced hospital length of stay. It is doubtful that PPS has led to reduced quality of care or access to care. Finally, while it is likely that the system has helped control Medicare's budget, it is unclear it has reduced costs per beneficiary.
12 Traditional economic descriptions depict regulation as a process in which the optimal policy is determined and imposed by an omniscient regulator. Alternative theories suggest that regulations result from political processes. If so, the regulations will reflect the preferences of the beneficiaries or "winners," moderated by the opposition of those who stand to lose if the regulations are imposed.

Discussion Questions

1 What is meant by market failure? What is the potential role of government in each instance of market failure found in the health care sector?
2 In what sense can information and redistribution be thought of as public goods? Explain whether private markets will oversupply or undersupply these goods.
3 What are some examples of government's providing health care? Of subsidizing the production or consumption of health care? Of providing insurance for health care? Of regulating health care markets?

4 Why is government needed to provide a public good? Under what circumstances might the voluntary contribution model, described in Figure 19.2, work reasonably well?
5 The standard monopoly–competition comparison describes the welfare loss. Develop arguments to support the view that in the real world:
   (a) the welfare loss is exaggerated;
   (b) the welfare loss is understated.
6 Discuss the nature of the negative externalities associated with the consumption of junk food. Discuss the pros and cons of a "junk food tax" and of regulations limiting portion sizes that can be served in restaurants.
7 Mandated health benefits have proliferated since 1970. Discuss the pros and cons of the ACA mandate requiring coverage of dependent children through to age 26.
8 What is meant by the tax subsidy of employer-paid health insurance? Explain why the subsidy very likely increases health care spending and thus the cost of such insurance. Give an example as to why a $5,000 health insurance benefit provides a greater monetary benefit to someone in the 33 percent marginal tax bracket than someone in the 15 percent marginal tax bracket.
9 Some economists propose a tax policy that would allow individuals who purchase their own insurance to deduct these costs as well as all out-of-pocket costs for health care from taxable income. Explain how this policy could help offset the bias toward more comprehensive insurance resulting from the preferential tax treatment of employer-paid insurance.
10 The pharmaceutical industry has been subject to considerable regulation in bringing drugs to the marketplace. Discuss the benefits and costs to society from such a policy.
11 How does prospective payment change the incentive to hospitals as compared to retrospective reimbursement? What predictions would one make due to the adoption of reimbursement based on DRGs?
12 Under Shleifer's theory of yardstick competition, why does the firm have an incentive to reduce its costs? If all firms respond by reducing their costs, will the payment rate also subsequently fall? (Students with training in game theory may wish to examine the elegant logic in the original article.)
13 What is the Medicare PPS program under DRGs? How has the Medicare PPS payment under DRGs affected hospital practices? Length of stay? Quality of care? Financial condition?
14 Does the Medicare PPS payment under DRGs reduce costs? Discuss.

Exercises

1 What is meant by the welfare loss of monopoly? Who bears this loss? Determine the regulated price in Figure 19.1 that will eliminate the welfare loss. Why will it be difficult in practice to adopt this solution?
2 Suppose that the public good in Figure 19.2 is associated with increasing costs of production (this would occur if *AC* is positively sloped). Will the two individuals be able to fund the optimal amount if each pays according to marginal benefits? Repeat if there are decreasing costs of production.
3 Assume a positively sloped, short-run supply curve in Figure 19.3 and a constant $5 per unit marginal external benefit. Show what happens with a $5 subsidy given to producers. Who gains the benefits of this policy?

4 Using Figure 19.3, explain how a $5 tax on those who are not inoculated will result in the efficient output. Why is this approach unlikely to work in practice?
5 Take a commodity, such as cigarettes or "junk food," associated with negative externalities. Assume that the marginal private cost of production and the marginal external cost per unit are both constant. Graph the actual and efficient quantities. Show how a tax on producers (or consumers) can result in the efficient output. Explain whether it is possible to attain efficiency with a subsidy.
6 Draw a graph showing the marginal revenue (MR) and marginal cost (MC) of increasing the tax rate (t) in the Peltzman model of regulation represented by equation (19.1). Place the tax rate on the horizontal axis and MR and MC on the vertical axis. Show the optimum tax rate.
7 In Figure 19.1, suppose the demand for the good was summarized by the equations:

$$P = 100 - 0.5Q$$
$$MR = 100 - Q$$

and that the marginal cost equals the average costs at $10 per unit.
(a) Calculate the optimum market quantity in a competitive market. (Hint: Set price equal to marginal cost.)
(b) Calculate the quantity brought to market by the monopolist.
(c) Calculate the monopolist's profit.
(d) Calculate the deadweight loss to society from the monopoly.
8 Suppose that Hospitals A through E have the following marginal costs for a given procedure:
Hospital A—$2,000
Hospital B—$2,200
Hospital C—$1,800
Hospital D—$2,700
Hospital E—$2,300
Calculate the yardstick price that would be assigned to each hospital. Which two hospitals will be assigned yardstick prices that do not cover their current marginal costs?
9 Here is a complex yardstick problem. A monopoly hospital faces the following demand curve

$$q = 400 - 10p$$

and the following marginal cost (with no fixed costs)

$$c = 22$$

(a) Calculate the profit-maximizing values of p^* and q^*, the maximized profit π^*, and the consumer surplus CS^*.
Suppose that the firm could reduce its costs according to the formula

$R = 40d^2$, where d = the original cost (here, 22) – the new (reduced) cost

A yardstick regulator assigns the hospital the following parameters:

Lump sum subsidy = 300;
Yardstick price = 20.

(b) Give the profit-maximizing condition for the yardstick regulation.
(c) Calculate the profit-maximizing values of p^* and q^*, cost reduction expense R^*, maximized profit π^*, and consumer surplus CS^*.

Notes

1 A \$5 subsidy *to the consumer* will shift the demand curve up by the \$5 so that the new demand mirrors *MSB* and passes through point *B*. The solution remains the same. Consumers buy quantity Q_2 at price P_1, paying P_2 net of the subsidy. It will generally be easier to administer commodity taxes and subsidies through producers than consumers.
2 Many students are familiar with the "Prisoner's Dilemma" (PD) game, as discussed in Box 14.1. PD is but one of many economic strategy games that can provide important insights into economic analyses.
3 The OECD provides comprehensive annual health care data for its members. For the United States, these data are somewhat different than those reported in its NHE accounts. At 46.6 Chile had a slightly lower public spending share than the United States, with The Netherlands topping the list at 87.6 percent (oecd.org/els/health-systems/health-expenditure.htm: Accessed February 18, 2016). See Paris (2010) for a comprehensive overview of health financing and other characteristics across most OECD countries.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 20

Social Insurance

In this chapter

- Social Insurance Policies and Social Programs
- Historical Roots of Social Insurance
- Affordable Care Act (ACA) of 2010
- Medicare and Medicaid in the United States
- Public Insurance and Health
- The Effects of Medicare and Medicaid
- Conclusions

Throughout this book, we have emphasized the role that markets can play in providing health care. We have also generally used economic efficiency, provided in theory by perfectly competitive markets, as a standard against which to judge the costs and benefits of policies. In this chapter, we address instead the issues that arise when a society considers providing for health care by offering health insurance, to some significant degree, at the public's expense. Such insurance programs provided through taxes or regulations are called social insurance programs. Having provided an overview of the rationales for social insurance in health care, we now turn to an examination of social insurance. We begin by considering the history of health care social insurance throughout the world, with emphasis on the United States. We then examine Medicare, Medicaid, and CHIP, and their effects, and we close with prospects and issues for the future.

Social Insurance Policies and Social Programs

Social insurance programs can be broken down into five categories. While not mutually exclusive, programs of the following five types exist in the United States and in other countries.

1 ***Poverty.*** Poor people lack purchasing power to buy the goods considered necessary to provide the minimal standards of decent life. Programs directed toward persons experiencing poverty involve either cash (more recently, debit cards) or goods "in kind," such as rent vouchers or food stamps.
2 ***Old Age.*** The elderly have attained a certain age, generally coinciding with retirement from active employment. Programs include income maintenance, such as Social Security, as well as services and considerations (such as assisted housing, Meals on Wheels, or transportation) that may address their generally decreased mobility.
3 ***Disability.*** The disabled have either temporary or permanent inability to work because of illness or work-related injuries. Programs generally provide cash benefits. Disability programs were among the earliest social insurance programs available.
4 ***Health.*** Programs cover illness or well-care financing and/or provide facilities for various groups. In the United States, most programs have targeted children, the elderly, and/or the poor, with the government financing the individual's health care either entirely or in part. In other countries, governments have more direct involvement in the financing and delivery of health services for larger segments (or all) of the population.
5 ***Unemployment.*** The unemployed receive assistance due to a temporary loss of work. While unemployment-related programs generally provide short-term cash benefits, in many countries, longer-term unemployment may lead into poverty-related welfare programs.

Analysts find several other definitions useful in considering social insurance programs. Some programs, termed *entitlements*, are available to all who qualify. Food assistance in the form of stamps or debit cards, for example, is available in the United States to all households whose incomes fall below specified levels, related in part to income, family size, and geographic location. Medicaid, or financed health care for the poor (discussed later in this chapter), is also an entitlement program for all who meet particular qualifications. In contrast, U.S. housing programs are not entitlements. Only limited numbers of subsidized units (on the supply side) or vouchers (on the demand side) are available.

Many programs are means-tested in that they are available only to individuals or households who meet certain income criteria. Households that receive aid for poverty-related

problems may lose some or all of the aid as their incomes increase. Such reductions in aid may have the unintended effects of discouraging efforts by low-income households to find jobs. For example, formulas that reduce poverty-related aid by $1 for each $1 earned on the job constitute taxes on job earnings at rates approaching 100 percent.

Finally, aid may take various forms. Programs often tie aid to the purchase of certain items, such as food or housing. This procedure ensures that the people use the aid to buy items that the legislators have deemed important. Under certain circumstances, however, it may be more economically efficient to provide a cash subsidy rather than one in kind.

Answer the question, "Which would you prefer, $100 in cash or $100 worth of physician care?" Most would answer "$100 in cash!" Alternatively, "Which would you prefer, $100 in cash or x dollars [less than $100] that you can spend any way you want?" If x is less than $100 (say $70) it shows that the recipient values the aid at less than $100, and society could help him or her out for less than $100. However, legislators and the voting public often seem to prefer subsidies *in kind* rather than in cash so that they can monitor or control the purchases of those receiving the subsidies. Food subsidy rules limit the purchase of "nonfood" items such as cigarettes or liquor, deemed undesirable by many, but they also prevent the purchase of laundry detergent or toilet tissue, which most would view to be desirable.

Program Features

We discuss certain common features to characterize health-related social insurance programs in the United States. The first three relate to receipt of care:

1 Contributions—taxes, deductibles, and coinsurance.
2 Benefits—how much, who is included, and what types of treatment are included.
3 Length of coverage.

The latter two describe the provision of care, as well as the political problems involved in initiating plans:

4 Means of reimbursement to providers.
5 Methods of determining payment levels to providers.

Although supported by government, most social insurance plans also impose costs on their recipients. Many are funded by tax collections, and care recipients are often taxpayers. In some cases, the taxes in question may be regressive. By definition, a regressive tax is one for which lower income people pay higher fractions of their incomes to the tax than do richer people. In contrast, a progressive tax is one in which lower-income people pay lower fractions of their incomes to the tax than do richer people.

In the United States tax payments into the Social Security program have always been somewhat regressive, a necessary condition for its 1935 passage. From the beginning, the tax has been a constant percentage of wage income, up to a ceiling at which the marginal tax rate becomes zero.[1] This means that on average, people in higher-income groups pay a smaller proportion of their income in payroll taxes. Both workers and their employers pay 6.20 percent, and the maximum wage base was $118,500 in 2016. Since workers with wages of $150,000 pay the same amount, 6.20 percent of $118,500, or $7,347, the tax percentage for the more affluent is 4.90 percent, rather than 6.20 percent below the $118,500 cap. The effective tax rate obviously falls as wage income rises above $118,500. In contrast, the Medicare tax rate is now 1.45 percent of all wage income, a constant percentage.

To determine whether a social insurance program is redistributive—whether it in net causes a transfer of money from the rich to the poor—one must consider not just the tax payments made but also the benefits received. For example, data for the Social Security program in the United States have tended to show that Social Security is redistributive—that the poorer people tend to gain more in net than do the rich. However, an interesting line of research, as noted in Box 20.1 indicates that differential longevity among income groups may be undoing this result.

BOX 20.1

Increased Longevity Favors the Rich in Social Security

Social Security was designed to ensure that no workers go penniless in old age, and to redistribute resources from wealthier to poorer Americans. The Social Security website describes the original benefit structure, set to go into effect in 1942.

> The original Act provided for monthly retirement benefits, payable to persons 65 and older who were no longer working. The benefit formula was based on cumulative wages (earned since 1937) in covered employment (initially covering only about half the jobs in the country, which were in commerce or industry). Specifically, monthly benefits equaled **1/2** of 1 percent of the first $3,000 of cumulative wages, plus **1/12** of 1 percent of the next $42,000, plus **1/24** of 1 percent of the next $84,000. So, for example, someone who retired in January 1942 (when benefits were scheduled to begin) after earning a total of $6,000 during the 5-year period from 1937 to 1941 would receive a benefit equal to $17.50 a month. This can be thought of, loosely speaking, as a typical benefit because the average worker at the time earned about $100 a month (which totals $6,000 after 5 years). Thus, although the Social Security Act was enacted in the middle of the Great Depression, it originally envisioned relatively small benefits that were not payable for several years.
>
> This preceding benefit formula never became operational because of the amendments of 1939. Nevertheless, it does embody two important principles that still guide benefit payments today: [1.] benefits depend on work in covered employment, and [2.] benefits replace a higher proportion of earnings for low earners.

While the formula has changed over the years, it has maintained this goal. However, differential longevity has skewed progress toward the goal, as noted by *New York Times* writer Neil Irwin.

Citing a study from the Brookings Institute, Irwin notes that life expectancy for the bottom 10 percent of male wage earners turning 66 in 2016 has risen 0.7 years compared with what was expected for their low-income counterparts 30 years ago. In contrast, for the top 10 percent of male wage earners, life span rose 8.1 years in the same period. This implies 7.4 years of additional Social Security benefits for the affluent.

Because recent net outflows from Social Security have caused some to fear that the funds will run out of money, many fiscal centrists and conservatives have proposed

to increase the full retirement age. (As of 2016, it is currently 66 and on track to rise to 67, though retirees can take a reduced level of benefits at age 62 and higher levels up to age 70.)

However, notes Irwin, the life span differential suggests that such a change would fall heavily on the backs of the poor because increasing the retirement age would bring them even closer to their (lower) expected age of death.

Sources: www.nytimes.com/2016/04/24/upshot/rich-people-are-living-longer-thats-tilting-social-security-in-their-favor.html?_r=0, "People are Living Longer: That's Tilting Social Security in Their Favor," accessed April 28, 2016.

www.brookings.edu/~/media/Research/Files/Reports/2016/01/life-expectancy-gaps-promise-social-security/BosworthBurtlessZhang_retirementinequalitylongevity_012815.pdf, "Later Retirement, Inequality in Old Age, and the Growing Gap in Longevity between Rich and Poor," accessed April 28, 2016.

www.ssa.gov/policy/docs/v66n1p1.html#mt5, Social Security: A Program and Policy History, accessed April 28, 2016.

In addition to tax payments, eligible recipients must often pay deductibles or coinsurance. Analysts also consider time costs for paperwork or waiting time for appointments or treatment.

Benefit levels and length of coverage are similar to the workings of private insurance. Given the equity considerations of social insurance, political considerations may affect both. Determining who and which treatments are covered is also important. Coverage of individuals may involve children and spouses. Coverage of treatments may mandate coverage of certain diseases (i.e., end-stage kidney disease) and exclude others (i.e., optometric or chiropractic services).

Health-related social insurance also has supply-related characteristics. In some programs patients may pay directly for expenses and then be reimbursed. In others, government may pay providers directly. In some countries, all physicians who participate in the national health care program are government employees.

Historical Roots of Social Insurance

European Beginnings

Prior to the passage of the ACA, the United States was the only industrialized country lacking a comprehensive health-related social insurance system. Historians date the pioneering legislation for a system of compulsory national health insurance to Germany in 1883. National health insurance spread to other European countries at the end of the nineteenth and the early part of the twentieth centuries.

The German system and the other European systems extended already-existing voluntary associations, often guild or mutual aid groups whose benefits to members included the pooling of insurable risks. The German laws of 1883 set up a highly decentralized program that covered workers in mining, transportation, construction, manufacturing, mechanical trades, and establishments using power machinery. Austria (1888), Hungary (1891), Sweden (1891), Denmark (1892), and Belgium (1894) followed. With Spain's adoption of a plan in 1929, nearly every European country had enacted health insurance laws. Some were compulsory, as in Germany, but other countries such as Belgium, Denmark, Sweden, and Switzerland, provided government subsidies to voluntary mutual insurance funds.

The United Kingdom established its first social health insurance system in 1911. That legislation helped prepare for the establishment of the British National Health Service in 1946, the most prominent example in the Western countries. Government provision was also common in Eastern European countries and the former Soviet Union, which began its system in 1926.

Early Experience in the United States

The United States came late to social insurance and to governmental health insurance in particular. In the late nineteenth and early twentieth centuries, Americans, like the Europeans, established voluntary group purchasing arrangements, and mutual benefit associations. However, in the United States, government did not take up the funding of these voluntary societies, compared with Germany in 1883 and the United Kingdom in 1911.

The major advance in U.S. social insurance occurred with the establishment of the Social Security program in 1935. Despite the social insurance thrust of the program and the reform-minded support for it, the legislation made concessions to political opposition to the New Deal, including the omission of governmental health insurance. The omission of health insurance from the Social Security Act was by no means the Act's only conservative feature. It relied on a regressive tax and provided no coverage to some of the very poor, such as farm laborers and domestics.

Proponents of compulsory health insurance plans were no more successful through the 1940s and 1950s. President Harry Truman proposed a single health insurance system that would include even those workers not covered by Social Security. During the public debate, opponents of compulsory health insurance called it "socialized medicine," a term that greatly weakened its support in the political climate of the Cold War. Truman won the 1948 election, but his success did not translate into a health insurance program.

The Establishment of Medicare and Medicaid

The social insurance debate in the United States has often ranged between those who believe in voluntarism and voluntary insurance, as won in negotiations between self-reliant industries and unions, and those who believe that only a compulsory insurance program would provide the insurance that was necessary for the larger population. The major social insurance programs for health care in the United States, Medicare and Medicaid, passed in 1965. President Lyndon Johnson had supported health care for the aged, and by winning a landslide victory in the 1964 election, he was able to push for these programs.

While Medicare and Medicaid have achieved widespread political support since the 1960s, fundamental concerns stem from their rising costs. To the present day, with growing elderly and poor populations as well as continual health care cost inflation, many observers perceived the total costs of the social insurance as having grown out of control. Politicians have sought to ensure that Medicare would have a sufficient trust fund to meet future needs, while balancing these needs against the increased taxes and payments for services necessary to support them.

The wider issue of social health insurance for the population as a whole had not yet been decided. The experience of the 1993 and 1994 Clinton health plan was an example. Bill Clinton won the 1992 election with a campaign promising health system reform. There had been signs of growing interest among the electorate in health system reform including social health insurance. The Clinton plan emerged during January through May 1993 through a large task force of government officials, health policy experts, congressional staffers, and others. With many perceiving the task force to be secretive and unresponsive to the public,

plan supporters were unable to develop a coalition of interests around clearly defined features of reform, while insurers lobbied and advertised against the plan. During 1994, the Clinton administration re-worked the plan to respond to critics but the public came to believe that it would entail considerable government bureaucracy, cost, and inefficiency. The plan was scrapped without a Congressional vote in late summer 1994. The opposition Republican Party gained control of both houses of Congress in that November's elections (for the first time in 40 years), ending that episode of reform.

The Affordable Care Act (ACA) of 2010

The 2008 U.S. Presidential election set the stage for potential health policy reform. With recurring worries about rising health care costs, and with numbers of the uninsured increasing due to the "Great Recession," the major Democratic candidates (Barack Obama and Hilary Clinton), and Republican candidate John McCain all spoke to the need for health care reform. Obama's election and increased Democratic majorities in the Congress led to passage in March 2010. While the details and the legislation will get more detailed treatment in Chapter 22, the Act as passed does the following:

- It requires most U.S. citizens and legal residents to have health insurance, the so-called *individual mandate*.
- It assesses a fee against employers with 50 or more full-time employees that do not offer coverage as a premium tax credit; this fee is $2,000 per full-time employee, excluding the first 30 employees from the assessment. It requires employers with more than 200 employees to enroll employees automatically into health insurance plans offered by the employer.
- It subsidizes state expansion of Medicaid to all non-Medicare eligible individuals under age 65 (children, pregnant women, parents, and adults without dependent children) with incomes up to 133 percent of the Federal Poverty Level (FPL) with a benchmark benefit package.
- It creates state-based Health Benefit Exchanges and Small Business Health Options Program (SHOP) Exchanges, administered by a governmental agency or nonprofit organization, through which individuals and small businesses with up to 100 employees can purchase qualified coverage. States that choose not to create their own exchanges can use federally-created exchanges.

The Act includes many other important provisions such as the imposition of a "Cadillac" tax on high cost employer-provided policies, and the elimination of underwriting policies that enable insurers to deny individuals insurance or charge them higher premiums based on pre-existing conditions. Some provisions of the Act took effect on enactment of the legislation; other provisions were to be phased in through 2018.

Political opposition arose immediately, focusing on the individual mandate, as well as concerns about the projected costs. However, implementation has proceeded (bolstered by favorable Supreme Court decisions in 2012 and 2014), and by 2020 the ACA has been expected to insure at least 32 million of the 50 million currently uninsured (noncitizens and illegal immigrants will not be allowed to participate). This represents the biggest expansion of U.S. health policy since the passage of Medicare and Medicaid in 1965. We discuss the ACA in much more detail in Chapter 22.

Medicare and Medicaid in the United States

While ACA will change large parts of the federal health care policy presence, Medicare and Medicaid continue to form the foundation. Medicare is a national program that primarily provides compulsory hospital insurance to the elderly plus optional medical coverage to which nearly all elderly subscribe. In contrast, Medicaid is operated by the states with matching federal dollars. It primarily provides health care coverage to people who are poor.[2]

Medicare

Medicare traditionally consisted of Hospital Insurance (HI), also known as Part A, and Supplementary Medical Insurance (SMI), also known as Part B. A third part of Medicare, sometimes known as Part C, the Medicare Advantage program, was established as the Medicare + Choice program by the Balanced Budget Act (BBA) of 1997 (Public Law 105–33) and subsequently renamed and modified by the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003 (Public Law 108–173). The Medicare Advantage program expanded beneficiaries' options for participation in private sector health care plans. In 2006 the MMA established a new prescription drug benefit, also known as Part D.

When Medicare began on July 1, 1966, approximately 19 million people enrolled. By 2014, nearly 55 million people were enrolled in one or both of Parts A and B of the Medicare program, and almost 12 million of them had chosen to participate in a Medicare Advantage plan. Figure 20.1 displays the growth of the Medicare program since its inception.

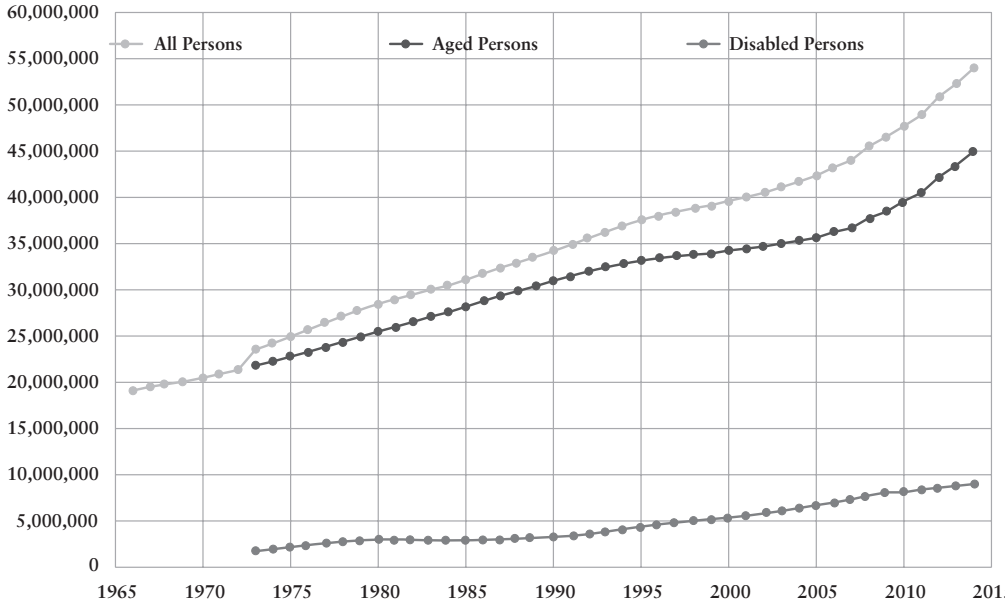


Figure 20.1 Medicare Enrollment by Year, 1966–2014

Source: Graph developed by authors through data from Centers for Medicare and Medicaid Services data compendium.

MEDICARE COVERAGE Medicare Part A generally goes automatically to persons age 65 and over who are entitled to Social Security or Railroad Retirement Board benefits. Similarly, those who have received such benefits based on a physical disability for at least 24 months also are entitled to Part A benefits.

Part A coverage includes:

- Inpatient hospital care coverage, requiring an initial deductible payment, plus copayments for all hospital days following day 60 within a benefit period.
- Skilled nursing facility (SNF) care—generally covered by Part A only if within 30 days of a hospitalization of three or more days and certified as medically necessary.
- Home Health Agency (HHA) care, including care provided by a home health aide.
- Hospice care, provided to those terminally ill persons with a life expectancy of six months or less and who elect to forgo standard Medicare benefits, receiving only hospice care SMI benefits (Parts B and D)—available to almost all resident citizens age 65 and over.

Part B coverage is optional and requires payment of a monthly premium. Part B covers

- Physicians' and surgeons' services (in both hospital and nonhospital settings).
- Some covered services furnished by chiropractors, podiatrists, dentists, and optometrists.
- Services in an emergency room or outpatient clinic, including same-day surgery, and ambulance services.

Part B also covers other services including clinical laboratory tests, X-rays, diagnostic tests, ambulance services, and blood that are not supplied by Part A. Almost all persons entitled to Part A also choose to enroll in Part B.

Medicare does not cover everything. Noncovered services include long-term nursing care, custodial care, and certain other health care needs, such as dentures and dental care, eyeglasses, and hearing aids. These services are not a part of the Medicare program unless they are a part of a private health plan under the Medicare Advantage program.

Part C—Medicare Managed Care

Medicare Advantage provides the option for beneficiaries to receive their Part A and Part B Medicare benefits through private health plans, mainly health maintenance organizations (HMOs), as an alternative to the federally administered traditional Medicare program. Medicare Advantage must pay for everything that Part A and Part B do, and some provide prescription drug care. In the early part of the 2000s, enhancements to the funding formula resulted in Medicare's paying private plans 14 percent more per enrollee than the cost of care for beneficiaries in traditional Medicare in 2009. The ACA of 2010 reduced federal payments to Medicare Advantage plans over time, bringing them closer to the average costs of care under the traditional Medicare program. It also exerted more control, providing for new bonus payments to plans based on quality ratings, and required plans to maintain a medical loss ratio of at least 85 percent, restricting the share of premiums that Medicare Advantage plans could use for administrative expenses and profits.

Medicare recipients over age 65 have remained one of the last bastions of fee-for-service coverage. In 2015, the majority of the 55 million people on Medicare received coverage through traditional Medicare, with 31 percent enrolled in a Medicare Advantage plan. Figure 20.2 illustrates that while the vast majority of those under age 65 have moved to managed care, even in 2015 only five states had at least 40 percent of eligible beneficiaries in

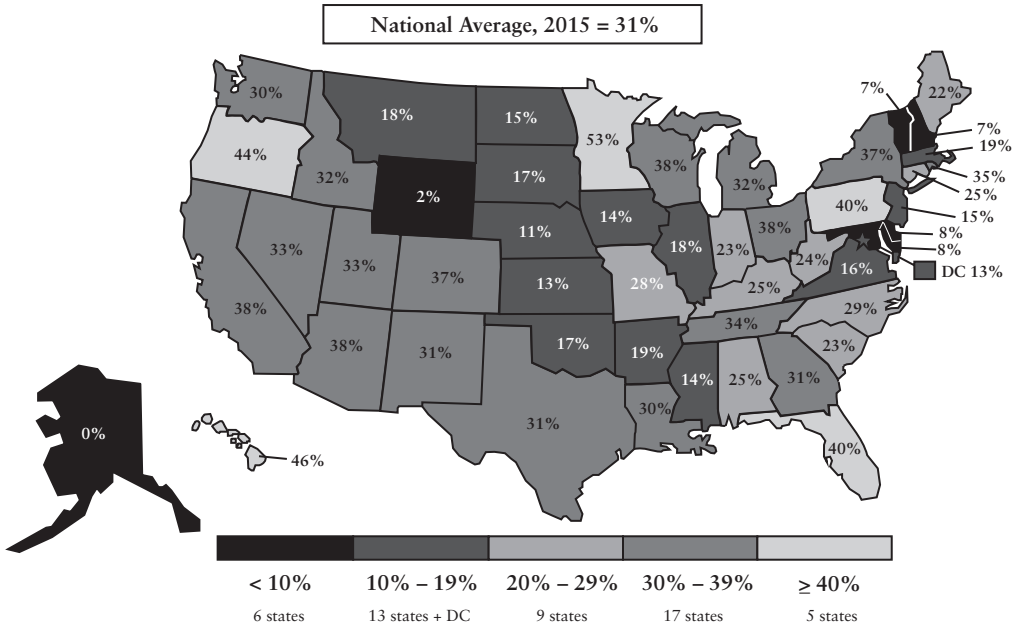


Figure 20.2 Share of Medicare Beneficiaries Enrolled in Medicare Advantage Plans, by State, 2015

Note: Includes MSAs, cost plans, and demonstrations. Includes Special Needs Plans as well as other Medicare Advantage plans.

Source: Jacobson et al. (2015). This information was reprinted with permission from the Henry J. Kaiser Family Foundation. The Kaiser Family Foundation is a nonprofit private operating foundation, based in Menlo Park, California, dedicated to producing and communicating the best possible information, research, and analysis on health issues.

Medicare Advantage plans, and only Minnesota had topped 50 percent. The number of beneficiaries enrolled in Medicare Advantage private plans has more than tripled from 5.3 million in 2004 to 16.8 million in 2015 (source: http://kff.org/medicare/fact-sheet/medicare-advantage/, accessed February 29, 2016).

Part D—Prescription Drug Insurance

As recently as 2005, one-third of Medicare's 43 million elderly beneficiaries had no prescription drug coverage, often for critical and expensive drugs (Schneeweiss and colleagues, 2009). Some seniors faced the risk of spending large portions of their incomes on essential medications, economizing on their use (such as cutting pills in half), or going without altogether. In a study of the previously uninsured, Schneeweiss and colleagues, seeking to typify the uninsured population, found that out of 1.5 million patients age 65 and older identified in three pharmacy chains, 202,548 (13.7 percent) had no drug coverage from any source throughout 2005. The mean age of the primary study population was 77.4 years, and two-thirds were women. Seventy-one percent of the patients had used four or more different

medications in the six months before Part D, and sizable fractions had a Chronic Disease Score (an aggregate comorbidity measure based on current medication use) of 4 or higher (30 percent), or used antidiabetic drugs (10 percent) or nitrates (8 percent).

Beginning in 2006, Part D provided subsidized access to prescription drug insurance coverage on a voluntary basis upon payment of a premium, to individuals entitled to Part A or Part B, with premium and cost-sharing subsidies for low-income enrollees. Part D coverage has included most FDA-approved prescription drugs and biologicals. For an additional premium, plans might also offer supplemental coverage exceeding the value of basic coverage. To encourage employer and union plans to continue to offer prescription drug coverage to Medicare retirees, Part D provides for certain subsidies to those plans that meet specific criteria. The coverage is provided privately, so different plans vary, based on their promised coverage.

The general goal of Part D has been to cover relatively small drug expenditures, and to guard against catastrophically large ones. Figure 20.3 presents the features of a typical Part D coverage in 2016. Annual premiums varied by plan, but consider a typical moderate coverage at $50 per month or $600 per year. In 2016, there was a $360 annual deductible. After the deductible, Part D covers 75 percent of all incremental expenditures up to $2,310.

A controversial feature of Part D has been the so-called "doughnut hole."[3] As noted in the shaded part of Figure 20.3, charges above $3,310 into the $7,000 range have been subject to very high copayment rates, originally 100 percent. This means that after an initial subsidy, the enrollee would have to pay dollar per dollar for several thousands of dollars of drugs. However, above $7,000 dollars, Part D almost fully indemnified its recipients.

The 2016 Part D plan has a $4,850 catastrophic threshold. Upon having spent $4,850 on drugs (in addition to the monthly premiums), enrollees have to spend approximately 5 percent out-of-pocket on expenditures past that threshold.

The initial Part D financing plan left open conjecture as to who would participate. When the program began, the break-even point, where the participant was no worse off than not participating (that is, premium + out-of-pocket expenditures equal to total charges), occurred

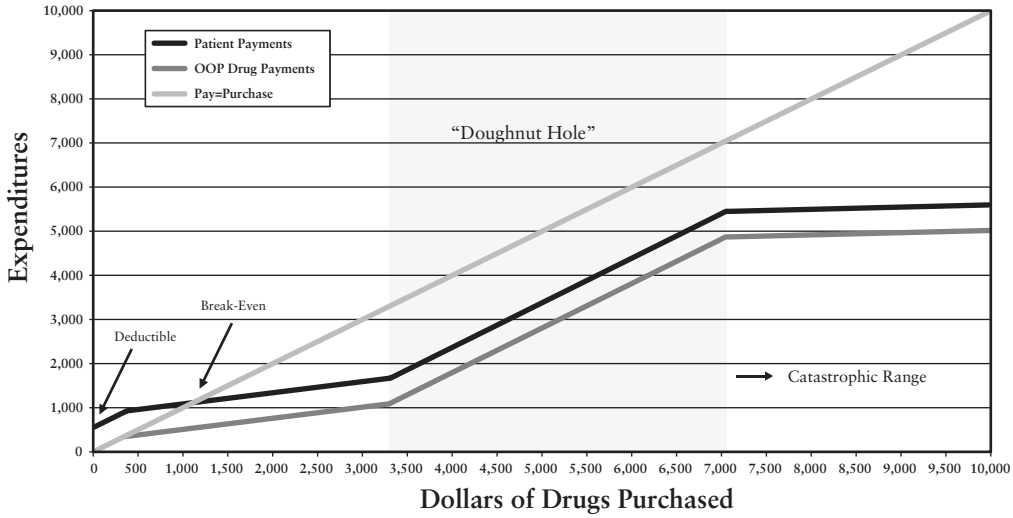


Figure 20.3 Medicare Part D Prescription Drug Benefit, 2016

at a value of $850 per year (about $70 per month).[4] While Part D initially appeared generous for people with small (initially up to about $210 per month) or large levels of expenditures (initially $425 per month, now over $500 per month), participants in the middle would derive little or no additional benefit as charges rose.

Levy and Weir (2010) provide an excellent evaluation of the impact of the Part D Benefit. Their "before-and-after" study framework allows them to examine the impact of an intervention (that is, the initiation of Part D) on a group seen prior to the intervention. They analyze data from the 2002, 2004 (both before Part D), and 2006 (after Part D was initiated) Health and Retirement Study on senior citizens' take-up of Medicare Part D and an associated Social Security Administration Low-Income Subsidy to help the lower-income elderly pay for Part D.

They find that economic factors—specifically, demand for prescription drugs—drove the decision to enroll in Part D. For the most part, individuals who already had employer-sponsored coverage kept that coverage, as they should have. Take-up of Part D among those without previous (2004) drug coverage was high; about 50 to 60 percent of them had Part D coverage in 2006. Only 7 percent of senior citizens lacked drug coverage in 2006 compared with 24 percent in 2004. Many of those who remained without coverage in 2006 reported that they did not use prescribed medicines, and the majority had relatively low out-of-pocket spending.

When Part D began, many felt that the program was too complicated for the elderly to use, but Levy and Weir reported that the majority of those interviewed had little or no difficulty with the Part D enrollment decision and were confident that they made the right decision. For the most part, then, despite the complexity of the program, Medicare beneficiaries were able to make economically rational decisions in which they had confidence. This too is not surprising. For those without any coverage, the decision to buy Part D coverage was hardly a marginal decision—almost any type of Part D plan was better than nothing.

The coverage gap, or doughnut hole, did induce some substitution behaviors. Hoadley and colleagues (2007) examined nationwide retail pharmacy claims data for 2007 and found that about 74 percent of the enrollees (excluding those enrollees who received low-income subsidies and nonusers) did not reach the coverage gap, about 22 percent remained in the coverage gap, and about 4 percent reached the catastrophic coverage level. Among eight drug classes, the majority of enrollees who reached the coverage gap made no detectable change in their medication use for the drug (or drugs) they were taking within the class when they reached the gap.[5] However, averaged across the eight classes, 20 percent of those who reached the gap made some change in their use of drugs within the selected class, while others may have stopped taking a drug in another class to continue taking medication in the studied class. In particular:

- 15 percent stopped taking their medication within the particular class,
- 5 percent switched to another medication (most often a generic drug) in the same class, and
- 1 percent reduced the number of separate medications they were taking in the class.

Has Part D impacted total health expenditures and/or improved recipients' health? Kaestner and Kahn (2012) found Medicare Part D significantly reduced socioeconomic and geographic disparities in elderly prescription drug insurance. Gaining prescription drug insurance through Medicare Part D related to a 30 percent increase in the number of annual prescriptions and a 40 percent increase in expenditures on prescription drugs for both the

general population of the elderly and the elderly in poorer health. The researchers found "little evidence" that prescription drug insurance was strongly associated with the use of outpatient and inpatient services.

In a follow-up study, Kaestner, Long, and Alexander (2014) examine whether obtaining prescription drug insurance through the Medicare Part D program affected hospital admissions, expenditures associated with those admissions, and mortality. Results indicate that obtaining prescription drug insurance through Medicare Part D was associated with an 8 percent decrease in the number of hospital admissions, a 7 percent decrease in Medicare expenditures, a 12 percent decrease in total resource use, and no significant change in mortality. These data allow the authors to estimate a total "offset" of $1.5 billion per year, or approximately 2.2 percent (the $1.5 billion of savings divided by $67.7 billion total state and federal expenditure) of the total annual cost of Medicare Part D.

In sum, eight years after it began (2014), Part D provided $86.4 billion in benefits to 39.2 million enrollees. The average benefit per enrollee exceeded $2,200.

MEDICARE PROGRAM FINANCING The Medicare Part A program is financed primarily through a mandatory payroll deduction (FICA tax). The FICA tax is 1.45 percent of earnings (paid by each employee and also by the employer) or 2.90 percent for self-employed persons. This tax is paid on all covered wages and self-employment income without limit.

The SMI trust fund differs fundamentally from the HI trust fund with regard to financing. SMI is now composed of two parts, Part B and Part D, each with its own separate account within the SMI trust fund. The financing for both parts of SMI is similar, in that both parts are primarily financed by beneficiary premiums and contributions from the general fund of the U.S. Treasury.

Financing for Part B comes from premium payments and contributions from the general fund of the U.S. Treasury. In 2016, new beneficiaries pay $121.80 per month. The patient premiums are indexed according to income, so those with incomes between $85,000 and $107,000 pay $170.50 per month, with the rate topping off at $389.80 per month for individual incomes over $214,000. Beneficiary premiums are generally set at a level that covers 25 percent of the average expenditures for aged beneficiaries. Therefore, the contributions from the general fund of the U.S. Treasury are the largest source of Part B income.

Similarly, Part D is financed primarily through premium payments and contributions from the Treasury general fund, with general fund contributions accounting for the largest source of Part D income, since beneficiary premiums are to represent, on average, 25.5 percent of the cost of standard coverage (as described in the next section). The premiums and general fund contributions for Part D are determined separately from those for Part B.

BENEFICIARY PAYMENT LIABILITIES Parts A and B beneficiaries must pay the charges not covered by Medicare and for various cost-sharing features of the plans. These liabilities may be paid by the beneficiary, by a third party, such as a private "Medigap" insurance policy purchased by the beneficiary, or by Medicaid, if the person is eligible. Medigap refers to private health insurance that, within limits, pays most of the health care service charges not covered by Parts A and B of Medicare.

For hospital care covered under Part A, the beneficiary's payment share includes a one-time deductible at the beginning of each benefit period ($1,260 in 2016). This covers the beneficiary's part of the first 60 days of each spell of inpatient hospital care. If continued inpatient care is needed beyond the 60 days, additional coinsurance payments ($322 per day in 2016) are required through the ninetieth day of a benefit period.

For Part B, the beneficiary's payment share includes the following: one annual deductible ($166 in 2016), the monthly premiums, the coinsurance payments for Part B services (usually

20 percent of the medically allowed charges), a deductible for blood, certain charges above the Medicare-allowed charge (for claims not on assignment), and payment for any services that are not covered by Medicare. For outpatient mental health treatment services, the beneficiary is liable for 50 percent of the approved charges.

PROVIDER PAYMENTS Before 1983, HI paid providers on a "reasonable cost" basis. Since 1983, HI has paid for most inpatient hospital services under prospective payment, or PPS. As discussed in Chapter 19, PPS pays a specific predetermined amount for each inpatient hospital stay, based on each stay's DRG classification. In some cases, the payment the hospital receives is less than its actual cost for providing the HI-covered inpatient hospital services for the stay; in other cases it is more. The hospital absorbs the loss or makes a profit. Certain payment adjustments exist for extraordinarily costly inpatient hospital stays, and payments for skilled nursing care and home health care, and rehabilitation and psychiatric care, are currently reimbursed on a reasonable cost basis, but prospective payment systems are expected in the near future.

Before 1992, under SMI, physicians were also paid on the basis of "reasonable charge," initially defined as the lowest of (1) the physician's actual charge, (2) the physician's customary charge, or (3) the prevailing charge for similar services in that locality. Changes beginning in 1992 defined "allowed charges" as the lesser of (1) the submitted charges or (2) the amount determined by a fee schedule based on a relative value scale (RVS). SMI reimburses most hospital outpatient services on a prospective payment system, with home health care reimbursed under the same prospective payment system as Part A.

Doctors or suppliers who agree to accept the Medicare-approved rate as payment in full ("take assignment") may not request any added payments, or "balance bill" (beyond the initial annual deductible and coinsurance) from the beneficiary or insurer. If providers do not take assignment, they will charge the beneficiary for the excess (which may be paid by Medigap insurance). Limits now exist on the excess that doctors or suppliers can charge. Physicians are "participating physicians" if they agree before the beginning of the year to accept assignment for all Medicare services they furnish during the year. Since Medicare beneficiaries may select their doctors, they have the option to choose those who participate.

MEDICARE SUMMARY The Medicare program covers 95 percent of our nation's aged population, as well as many people who are on Social Security because of disability. In 2014, Part A covered about 53 million enrollees with benefit payments of $264 billion, and Part B covered about 49 million enrollees with benefit payments of $256 billion. Parts A, B, C, and D together provided $618.7 billion dollars to 55.1 million enrollees. Administrative costs for both Parts A and B were $6.7 billion, or approximately 1.5 percent of disbursements.

Medicaid

Medicaid, referring to Title XIX of the Social Security Act, is a federal-state matching entitlement program paying for medical assistance for certain vulnerable and needy individuals and families with low incomes and resources. Medicaid is the largest source of funding for medical and health-related services for America's poorest people. In 2014, it provided health care assistance to more than 65.9 million persons. Expenditures for fiscal year 2014 totaled $495.8 billion.

Gruber (2002) described Medicaid as four public insurance programs in one. The first provides coverage of most medical expenses for low-income women and children families. The second provides public insurance for the portions of medical expenditures not covered by Medicare for the low-income elderly, and the third covers most medical expenses for the

low-income disabled. The fourth pays the nursing home expenditures of many of the institutionalized elderly.

The 2010 ACA accorded a new primacy to the Medicaid program. States would be given 100 percent federal financing for those made newly eligible for Medicaid under the ACA. The grant would fall to 95 percent in 2017, 94 percent in 2018, 93 percent in 2019, and then 90 percent in 2020 and beyond. We save the details for Chapter 22, but this major expansion provides a "fifth" insurance plan in Gruber's taxonomy.

Under Medicaid, each state, within broad national guidelines established by federal statutes, regulations, and policies, (1) establishes its own eligibility standards; (2) determines the type, amount, duration, and scope of services; (3) sets the rate of payment for services; and (4) administers its own program. Medicaid policies for eligibility, services, and payment vary considerably even among similar-sized and/or adjacent states and the services provided by one state may differ considerably in amount, duration, or scope from services provided in a neighboring state.

Medicaid Eligibility

The policymakers did not design Medicaid to provide medical assistance for *all* poor persons. Even under the broadest provisions of the 1965 federal statute, it may exclude some very poor persons unless they are in one of the designated groups. Low income is only one test for Medicaid eligibility for those within these groups; potential recipients' resources also are tested against threshold levels (as determined by each state within federal guidelines).

To be eligible for federal funds, states must provide Medicaid coverage for certain individuals who receive federally assisted income-maintenance payments, as well as for related groups not receiving cash payments. Although there is a long list, Medicaid "categorically needy" eligibility groups for which federal matching funds are provided to states include:

- Low-income families with children.
- Children under age six and pregnant women whose family income is at or below 133 percent of the federal poverty level (FPL).
- All children born after September 30, 1983, who are under age 19, in families with incomes at or below the FPL.
- "Dual eligible" Medicare beneficiaries.

Outside of these categories, however, states have had broad discretion in determining which groups their Medicaid programs will cover and the financial criteria for Medicaid eligibility.

Medicaid is a cost-sharing partnership between the federal government and the states. The federal government pays a share of the medical assistance expenditures under each state's Medicaid program. That share, known as the Federal Medical Assistance Percentage (FMAP), is determined annually by a formula that compares the state's average per capita income level with the national income average. States with higher per capita income levels are reimbursed smaller shares of their costs. By law, the FMAP cannot be lower than 50 percent or higher than 83 percent. In 2016, the FMAPs varied from 50 percent (13 states) to 74.2 percent (Mississippi), with the median federal share among all states being 55.5 percent.

This means that a state with an FMAP of 50 percent is matched 50 cents for every 50 cents that it contributes. Hence that state is paying at a rate of one-half ($0.50/$1.00) of the actual price. In contrast, a state with an FMAP of 74.7 percent is matched 74.2 cents for every 50 cents it contributes, giving a rate of 40.3 percent ($0.50/$1.242) of the actual price.

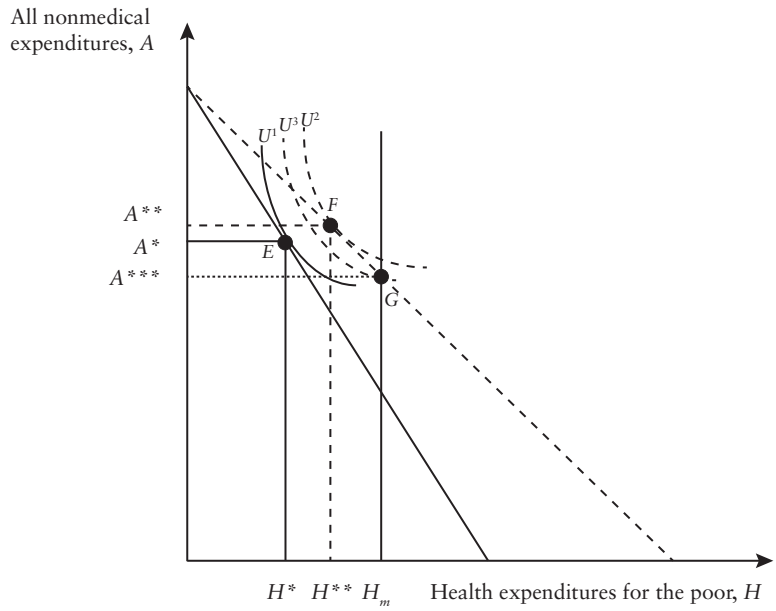


Figure 20.4 The Impact of Medicaid Cost Sharing on State Expenditures

Figure 20.4 shows how this formula can benefit individual states. Suppose that a state previously provided health services expenditure level H^* for the poor, and expenditure level A^* for everything else, at point E, leading to utility level U^1. These expenditure patterns presumably reflected the preferences of the public for taxing themselves to spend for various items for their residents. The –1.0 slope of the budget constraint reflects the relative costs of shifting one dollar *away* from one category *to* another. The Medicaid cost share reduces the relative cost of health care for the poor, thus rotating out the x-axis of the diagram, as shown by the dashed line. With a 1:1 match (50 cent match for every 50 cent expenditure), the slope of the curve rotates from –1.0 to –0.5. If the state faces no constraints as to how it uses the aid, it might be able to increase expenditures on both A and H and get to point F, and utility level U^2.

However, we should analyze Medicaid's requirements as constraints that require states to provide (and tax their residents for) a mandated expenditure level H_m of health services which exceeds H^{**}. Medicaid is thus not a "block grant" that the states can use as they wish; states must provide their own shares of expenses through taxes, and provide services to specified groups of the needy in order to participate. The requirement that states provide stipulated levels of services in order to receive Medicaid funding constrains state behavior, and may reduce the utility of the representative voter. This is utility level U^3 at point G. Although Medicaid program participation certainly increases the utility of the Medicaid beneficiaries, program mandates reduce utility level from level U^2 at point F. Nonetheless, the fact that all states choose to participate in regular Medicaid with its cost-sharing and mandated benefits indicates the overall desirability of this program to the states' residents. The refusal of many U.S. states to participate in the far more generous ACA matching can only be explained by strong preferences of governing majorities to withhold services from less affluent minorities.

THE SCOPE AND DURATION OF MEDICAID SERVICES The Medicaid program allows considerable flexibility within the states' Medicaid plans (see Box 20.2 for a particular example related to Oregon). Because the states do vary, analysts can compare state programs to determine how differing program features might work. However, a state's Medicaid program must offer medical assistance for certain basic services to most categorically needy populations, including inpatient hospital services, outpatient hospital services, prenatal care, vaccines for children, physician services, nursing facility services for persons age 21 or older, and family planning services and supplies.

Within broad federal guidelines and certain limitations, states determine the amount and duration of services offered under their Medicaid programs. States may limit, for example, the number of days of hospital care or the number of physician visits covered. States must provide sufficient levels of services to achieve the purpose of the benefits, and benefit limits may not discriminate among beneficiaries based on medical diagnosis or condition.

BOX 20.2

Oregon Medicaid's Doctor-Assisted Suicide—18 Years Later

Perhaps nowhere is the state-level autonomy in the U.S. Medicaid system more apparent than in the items various states choose to cover. In late February 1998, the state of Oregon's Health Services Commission voted 10–1 to include doctor-assisted suicide on the list of "treatments" covered for Medicaid patients, reported Peter Steinfels of the *New York Times*. This decision joined doctor-assisted suicide to other forms of "comfort care" for any "terminal illness, regardless of diagnosis." Residents of Oregon had voted twice, in 1994 and again in 1997, to legalize doctor-assisted suicide but neither vote had dealt with the public financing of the procedure.

Dr. Alan Bates, who headed the commission, acknowledged the divisive nature of the decision. He noted, however, that if dying people with private insurance could pay for medical help in taking their own lives, why should poor people not have the same opportunity?

In 2002, U.S. Attorney General John Ashcroft challenged Oregon's practices. After numerous appeals, the U.S. Supreme Court, in a 6–3 vote in January 2006, ruled that a federal drug law could not be used to prosecute Oregon doctors who prescribed overdoses intended to facilitate the deaths of terminally ill patients. For the majority, Justice Anthony Kennedy wrote, "[The Attorney General] is not authorized to make a rule declaring illegitimate a medical standard for care and treatment of patients that is specifically authorized under state law."

In the 18 years subsequent to the 1998 law (as of January 2016), 1,545 Oregon residents received prescriptions for lethal medications under the Oregon Death With Dignity Act (DWDA). Nine hundred and ninety-one (991) of them died from ingesting those medications.

Sources: Steinfels, Peter, "Oregon Medicaid's Doctor-Assisted Suicide," *New York Times*, March 7, 1998, National/Metro Section; http://public.health.oregon.gov/ProviderPartnerResources/EvaluationResearch/ DeathwithDignityAct/Documents/year18.pdf, accessed March 15, 2016.

PAYMENT FOR MEDICAID SERVICES Under Medicaid, states may pay health care providers directly on a fee-for-service basis or through various prepayment arrangements, such as HMOs. Each state has broad discretion in determining the payment methodology and payment rate for services. Generally, payment rates must be sufficient to enlist enough providers so that covered services are available at least to the extent that comparable care and services are available to the general population within that geographic area. Providers participating in Medicaid must accept Medicaid payment rates as payment in full. States must make additional payments to qualified hospitals that provide inpatient services to a disproportionate number of Medicaid beneficiaries and/or to other low-income or uninsured persons under what is known as the "disproportionate share hospital" (DSH) adjustment.

States may impose nominal deductibles, coinsurance, or copayments on some Medicaid recipients for certain services, but pregnant women and children under age 18 cannot be required to share costs. All Medicaid recipients must be exempt from copayments for emergency services and family planning services.

MEDICAID SUMMARY Medicaid started as a medical care extension of federally funded programs providing cash income assistance for the poor, with an emphasis on dependent children and their mothers, the disabled, and the elderly. Legislation in the late 1980s assured Medicaid coverage to an expanded number of low-income pregnant women, poor children, and some Medicare beneficiaries who are not eligible for any cash assistance program. Legislative changes also focused on increased access, better quality of care, specific benefits, enhanced outreach programs, and fewer limits on services. With the passage of the 2010 Affordable Care Act, Medicaid became a foundation of health insurance expansion, although some states have not chosen to participate.

Most Medicaid recipients require relatively small average expenditures each year. Data for 2011 showed mean Medicaid payments for all children of about $2,463 per child and for all adults of $3,427. This was much lower than that for aged ($13,249) and disabled beneficiaries ($16,643), reflecting the differing health status and use of services by the members of these groups.[6]

In 2013 Medicaid paid for 51 percent of long-term services and supports. With the elderly or disabled percentage of the population increasing faster than the younger groups, the need for long-term care is expected to increase.

The Medicaid–Medicare Relationship

The Medicare and Medicaid programs work jointly for many beneficiaries, called "dual eligibles." During 2013, more than 10.7 million Americans were enrolled in both the Medicare and Medicaid programs, a 24 percent increase from 2006. Two-thirds of this population were low-income elderly individuals, and one-third were individuals who were under age 65 and had disabilities. About 43 percent of Medicare–Medicaid enrollees had a Medicare-qualifying disability, compared to 12 percent of Medicare-only beneficiaries. Medicare–Medicaid enrollees have had a higher prevalence of many conditions (including, but not limited to, diabetes, pulmonary disease, stroke, Alzheimer's disease, and mental illness) than their Medicare-only and Medicaid-only peers. Medicare–Medicaid enrollees' health costs were four times greater than those of all other people with Medicare.[7]

For those eligible for full Medicaid coverage, the Medicare health care coverage is supplemented by services available under their state's Medicaid program, according to eligibility category. Additional services may include, for example, nursing facility care beyond the

100-day limit covered by Medicare, as well as eyeglasses and hearing aids. For those enrolled in both programs, Medicare pays first for services because Medicaid is always the "payer of last resort."

Medicare and Medicaid: Conflicting Incentives for Long-Term Care

The structures of Medicare and Medicaid can create conflicting incentives regarding dually eligible beneficiaries, without coordination of their care. Both programs seek to limit their own costs, but neither has an incentive to take responsibility for the management or quality of care.

David Grabowski (2007) explains that Medicare beneficiaries who meet Medicaid's (low) income and resource eligibility standards may become dually eligible (for both programs). Under federal rules, most states are required to offer Medicaid coverage to recipients of the Supplemental Security Income (SSI) program. However, Medicaid programs cover elderly people who have incomes up to 100 percent of the federal poverty level and assets that do not exceed the SSI threshold. The states adopted two broad sets of rules that expand income-related eligibility:

- "medically needy" programs, and
- special income rules.

If individuals' incomes exceed the state's income test, medically needy programs subtract medical and long-term care expenses from their incomes in calculating Medicaid eligibility. Other special income rules for people in nursing homes and in home- and community-based services (HCBS) waiver programs extend eligibility up to 300 percent of the SSI income limit.

Both Medicare or Medicaid may have the incentive to shift costs to the other. According to the rules, Medicare is the primary payer for dual eligibles' hospital, physician, and other acute medical care; Medicaid (according to the states' discretions) can choose to pay the often considerable Medicare copayments for the dual eligibles. If the states seek to reduce their Medicaid expenditures, they may restrict their cost-sharing paying. This may result in less access and less treatment for beneficiaries in states with more restrictive policies.

The adverse incentives can also go in the other direction. Most analysts believe that Medicare's 1983 adoption of hospital care DRGs led to patients being discharged "sicker and quicker." This change in payment contributed to the growth in Medicare-covered post-acute nursing home care in the years following prospective payment (Dalton and Howard, 2002).

The transfer of patients from the hospital to the nursing home also raises issues related to the coordination of care and the beneficiaries' health. Under the Medicare hospital prospective payment, discharge planners have more incentive to discharge patients as soon as (safely) possible but less incentive to consider the long-term cost and health implications of the initial discharge placement. With a high number of Medicare nursing home stays ultimately becoming Medicaid nursing home stays, care managers find it desirable that the "receiving" nursing home participate in Medicaid, even if Medicare finances the initial stay. Such placements would remove the need to transfer patients when their Medicare coverages end, thereby avoiding the adverse health consequences of transfers. Similarly, hospital discharge planners would ideally avoid transfers to nursing homes when adequate home care is available to support a community-based placement. This could improve patients' welfares and lower Medicaid's spending, but under the current Medicare payment system, discharge planners

are not rewarded for placing patients in the most appropriate setting, with little incentive to consider the long-term implications of the discharge placement for either the beneficiary's long-term health or Medicaid's budget.

What can be done about the conflicting incentives? The ACA established the Federal Coordinated Health Care Office (CHCO), known as the "office of the duals." This office works closely with the Center for Medicare and Medicaid Innovation to streamline care for dual eligibles. It addresses the cost-shifting and inconsistencies that can contribute to fragmentation of care, particularly as patients move back and forth from hospital, home, rehab, and long-term care—with some services and settings under Medicare's purview and some under Medicaid's.

Have coordination efforts proven successful? Jung et al. (2015) evaluate a program in which CMS partners with states to examine the financial and administrative alignment of Medicare and Medicaid services by integrating the benefits of both programs under a single entity. Twenty-six states were pursuing these programs, but the authors find little evidence to show program effectiveness.

They also examine an early demonstration for dual eligibles in Massachusetts of Senior Care Options (SCO), and its effect on rehospitalization. They find that SCO did not have a statistically significant effect on rehospitalization, an area where coordinated care would be expected to make a substantial difference. They observe that programs seeking to improve care for duals may need to consider not only the *structure* of benefits, but also the *specific interventions* used by plans and the characteristics of duals who are likely to enroll, so that participation can be appropriately gauged and services tailored accordingly.

Children's Health Insurance Program—CHIP

The State Children's Health Insurance Program, or SCHIP, was established in the Balanced Budget Act of 1997. Now called CHIP, and designed as a federal–state partnership, similar to Medicaid, it seeks to provide health insurance to children whose families earn too much money to participate in Medicaid, but not enough money to purchase private insurance. CHIP was the largest expansion of health insurance coverage for children since the initiation of Medicaid in the mid-1960s. The 2013 monthly "point in time" enrollment was 5.8 million, over twice as high as the 2000 enrollment.[8]

CHIP aims to provide coverage to "targeted low-income children." A "targeted low-income child" is one who resides in a family with income below 200 percent of the FPL or whose family has an income 50 percent higher than the state's Medicaid eligibility threshold. Some states have expanded CHIP eligibility beyond the 200 percent FPL limit, and others cover entire families and not just children.

If a state elects to establish an expanded Medicaid program using CHIP, the eligibility rules of Medicaid apply, and the services provided under CHIP mirror the Medicaid services provided by that state. Regardless of the type of health benefits coverage provided by a state, they must provide coverage for well-baby and well-child care, immunizations, and emergency services.

Public Insurance and Health

How does public insurance affect health? This depends in part on how effectively the public insurance programs reach their targeted populations. Janet Currie (2006) argues that countries with universal programs seek to maintain a minimum level of service for all individuals,

at a reasonable cost to government. In the United States, the goal before passage of the ACA was to maintain such a standard for selected groups of vulnerable or "deserving" individuals, such as children, the elderly, and the disabled. Targeting will never be perfect. Some who take up benefits will not "deserve" them, and some who are eligible for benefits will not take them up. If take-up by eligible individuals is low, the program may fail to reach its main goal of helping the targeted group. Take-up by ineligibles will divert government revenues from other productive uses.

Researchers have identified two categories of impediments to program take-up. The first is program *stigma*, meaning that some people are embarrassed or afraid to apply for programs, even though they might benefit greatly. Second, individuals face costs of learning about and applying for programs and these costs may deter some from using them. Moreover, the costs may be highest for precisely those individuals in greatest need, and in cases where the beneficiaries are young children or infirm adults, the costs may be borne by individuals other than the beneficiaries, such as parents or caretakers. These costs to would-be participants may be sufficiently large to prevent them from enrolling.

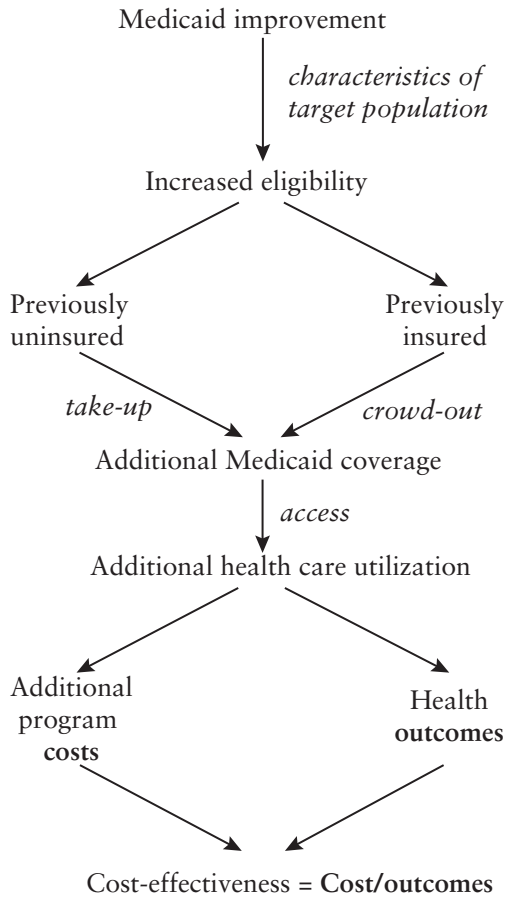


Figure 20.5 Analyzing the Impacts of Improved Medicaid Coverage
Source: Adopted from Gruber (2002).

Gruber (2002) traces the potential effects of a generic Medicaid improvement in Figure 20.5. The adoption or enhancement of a program such as Medicaid, depending on the population characteristics, will lead to increased eligibility of the poor or the young. Some may have been previously uninsured. To the extent that they find the public insurance attractive, they "take up" coverage. Some of those who were previously insured by other means may choose to substitute the public insurance. Researchers and policy analysts call this impact "crowd-out" since public insurance has replaced the private insurance. Either take-up or crowd-out will have some measurable impact on coverage.

Cutler and Gruber (1996) examine the economics of both take-up and crowd-out. Consider a household choosing between health insurance and all other goods, and assume that more generous plans offer a greater range of providers or cover a wider set of services. As noted in Figure 20.6, households valuing insurance highly (e.g., those demanding the highest quality providers) will exhibit utility function V_m (more services), and select point D. Those valuing insurance less highly will exhibit utility function V_l (less services), and select point E.

Suppose the government introduces free public insurance with generosity M. It may have a lower value relative to the private policies for a couple of reasons. Because of low Medicaid reimbursement rates, some providers may be reluctant to treat Medicaid patients. Some households may prefer to avoid public programs because of the stigma of being enrolled. Households cannot purchase a supplement to Medicaid; if they want higher insurance, they must return to the original budget constraint. Hence, the budget constraint is the kinked set of segments $ABMC$. Responding to the public coverage, people with low values of private insurance, such as those at point E, will enroll in the public sector, because utility level V_1' (passing through point M) is higher than V_1. Households with a high valuation of insurance

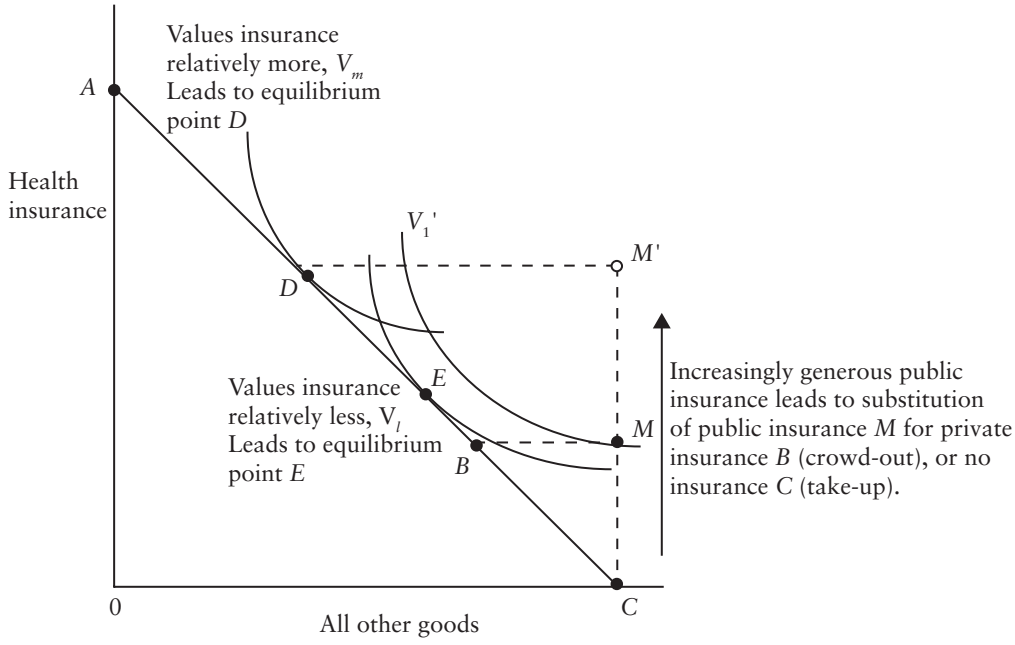


Figure 20.6 Analyzing the Economics of Take-Up and Crowd-Out

will retain their private insurance at point *D*. As the value of the public insurance (point *M*) rises (say to *M'*), the households are more likely to drop private insurance and enroll in Medicaid.

To the extent that the Medicaid coverage provides insurance where none has been available, one observes take-up; to the extent that it replaces existing insurance, one observes crowd-out. One may even see a household choosing less coverage (point *M* provides less generous coverage than existing point *E*) because it is free. Opponents of the Affordable Care Act have jumped on this possibility as a failure of the ACA, when in fact it is a feature of all programs that give recipients choices to purchase less of the specified good (for example, housing vouchers), for much less money, freeing up money for them to buy other items that they value more. The expansion of Medicaid through the Affordable Care Act has stimulated numerous analyses of the magnitudes of these effects, and we address them in more detail in Chapter 22.

Observers would expect increased coverage to affect health care utilization. Analysts have found that this impact depends on access to the health care, which may relate to the availability of providers and the distance, cost, or convenience of dealing with the providers. Increased utilization increases costs and presumably improves outcomes, which are typically measured in terms of reduced morbidity (illness) or mortality (death). The incremental cost per unit of outcome is often summarized in terms of costs per illness day prevented, or costs per death prevented—measures of the program's cost efficiency.

The impact of children's health programs has been substantial. While the percentage of children above 200 percent of the poverty level (the "not poor") stayed roughly constant from 1997 to 2010, the percentage of those below 200 percent, and particularly those below the poverty line, fell from well over 20 percent to less than 12 percent over the 13-year period. This is particularly notable given the increase in poverty that accompanied the hard economic times in the U.S. toward the end of the first decade of the twenty-first century.

The Effects of Medicare and Medicaid

Though we can be certain about the provisions of Medicare and Medicaid, we are necessarily less certain about their effects. We consider here a selection of findings on the effects of Medicare and Medicaid on:

- health care costs
- access to health care
- health status.

Costs and Inflation

The implementation of Medicare and Medicaid coincided with a considerable increase in health care costs in the United States. While health care costs had been rising before 1965, both in simple percentage terms and in comparison to the general rate of inflation, the hospital care inflation rate increased somewhat after the implementation of Medicare and Medicaid.

The expenditure levels of the two programs increased much more rapidly than most had expected. Figure 20.7 shows the pattern of expenditures over time. In monetary terms, percentage increases in expenditures on both programs were in double digits for many years in the 1990s, and the 2014 total of $1,127 billion is 8.4 times as great as the 1987 figure of $133 billion.

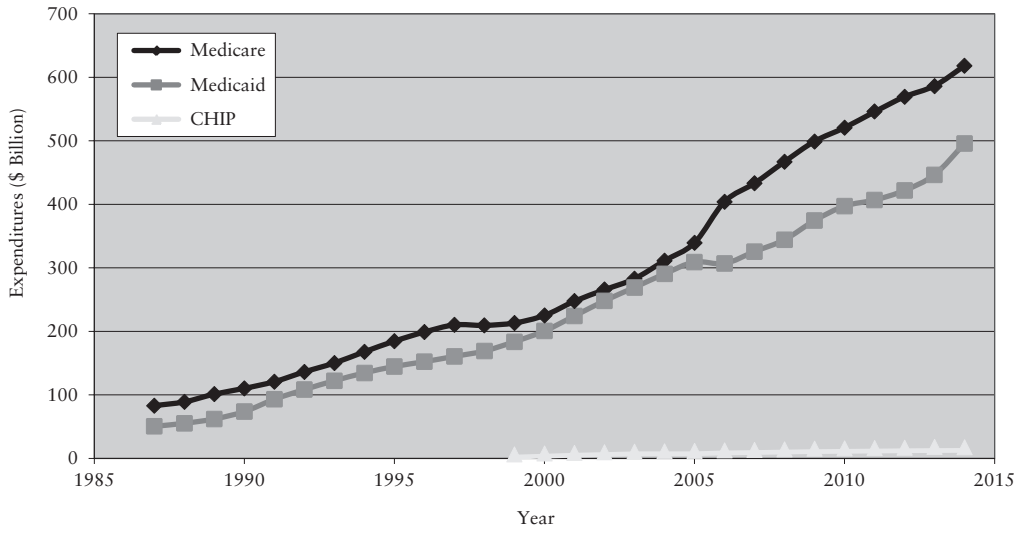


Figure 20.7 Total Expenditures for Medicare and Medicaid, 1987–2014

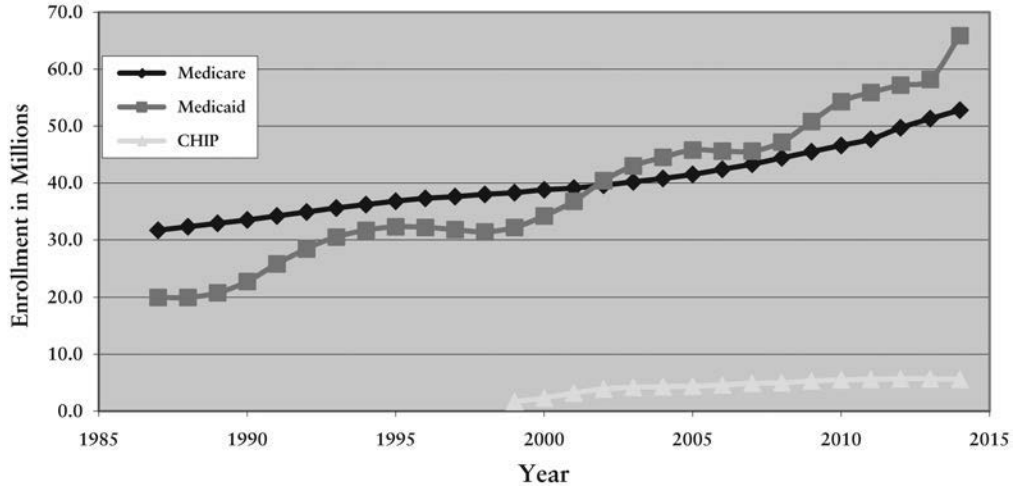


Figure 20.8 Enrollment for Medicare and Medicaid, 1987–2014

A large portion of the increase comes from enrollment, as noted in Figure 20.8. Medicare enrollment is largely predictable, and as the older population has increased, so has the enrollment, from 31.7 million in 1987 to 52.8 in 2014. Medicaid and CHIP have been more policy-driven. As Medicaid is one of the foundation blocks of the ACA, its enrollment has dramatically increased by a factor of more than three, from 20.0 million in 1987 to 65.9 million in 2014.

The inflation-adjusted expenditures per enrollee increased dramatically from 1987 through 2008, as noted in Figure 20.9. They have leveled off and even declined slightly over

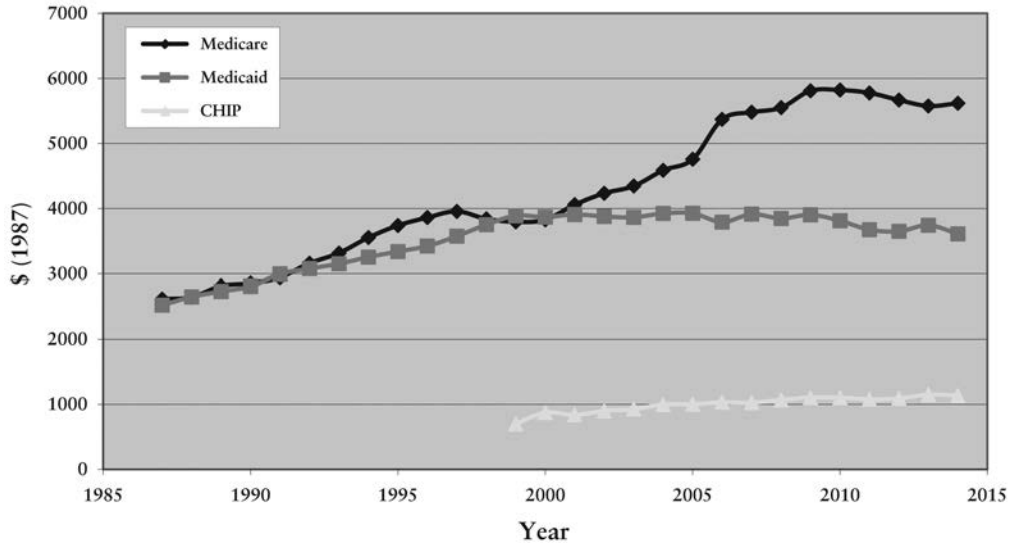


Figure 20.9 Real Expenditures per Enrollee, 1987–2014

the last half-dozen years, verifying that aggregate expenditure increases came almost entirely from enrollment increases. Policy analysts continue to debate whether this leveling off can be traced to the passage of the ACA in 2010, and whether it can be expected to continue.

WHY SPENDING HAS RISEN: INCREASED COVERAGE, TECHNOLOGICAL IMPROVEMENT, AND INCREASED INEFFICIENCY The increase in the eligible population covered by Medicare and Medicaid clearly helps explain why program expenditures have risen, but it does not fully account for the inflationary effects. Newhouse (1978) suggested three ways through which insurance programs, such as these, could affect prices and costs, even without growth in the population served.

First, Medicare and Medicaid both tended to increase the insurance coverage of the populations eligible. An increase in insurance expands the demand for care. Second, insurance coverage may induce technological improvements. If so, then the cost per unit of care may rise.

Finally, Newhouse proposed a third theory for the effect of insurance on costs and quantity used. This may be called the "increased inefficiency" theory. The idea is that when insurance covers a substantial portion of the health care bill, institutions, such as hospitals, have less incentive to control costs. It is not clear from this theory that the advent of Medicare and Medicaid, for example, would cause the level of inefficiency in hospitals to increase over time, but such a pattern is at least consistent with the theory.

THE EVIDENCE What do we know about the patterns of health expenditure inflation subsequent to the adoption of Medicare and Medicaid, and what do we know about the sources of this inflation? One approach partitions the observed rise in expenditures into its logical components: changes in population, in quantity per capita, and in the nature of services provided per visit or per admission.

Cutler and Meara (1997) find a dramatic change in medical spending over time, and disproportionately so for the very young (those younger than 1 year old) and the old (those age 65 or over). From 1963 through 1987, per-person spending on infants increased by

9.8 percent per year, and per-person spending on the elderly increased by 8.0 percent per year (compared to 4.7 percent per year for the others).

Moreover, they find that essentially all of the disproportionate growth of spending for the very young and the old was accounted for by high-cost users within those groups, and that a substantial amount of high-cost medical use is associated with the increasing technological capabilities of medicine. Among infants, high-cost users are premature babies with substantial respiratory or other acute conditions. For the elderly, high-cost users are generally patients with severe cardiovascular problems or cancer.

Finkelstein (2007) suggests that the impact of Medicare on hospital spending is substantially larger than what the existing evidence from individual-level changes in health insurance would have predicted. She argues that the introduction of Medicare was associated with an increase in the rate of adoption of then new medical technologies. A back of the envelope calculation based on the estimated impact of Medicare suggests that the overall spread of health insurance between 1950 and 1990 may explain at least 40 percent of the increase in real per capita health spending over this period.

Irrespective of the considerable costs, the predominant evidence seems to suggest that both Medicare and Medicaid have succeeded in addressing problems of access. That there was a change in health care use rates among the lower-income groups and the elderly following the beginning of Medicare and Medicaid is evident from the data.

Table 20.1 investigates two dimensions of health care access and utilization: the interval since the last physician contact and the number of hospital discharges per 1,000 people. In 1964, just before the passage of Medicare and Medicaid, 69.7 percent of those aged 65 and older had seen a physician within the past year. This was 4.2 percent higher than the general population and 8.1 percent higher than those aged 45 to 64. By 1990, those aged 65 and older were 11.4 and 12.7 percent more likely to have seen a physician within a year than the respective comparison groups.

Table 20.1 Indirect Impacts of Medicare and Medicaid

A. Age Comparisons

Interval since last physician contact

	Percent Less Than One Year		
	1964	*1990*	*1995*
Total	66.9	78.2	79.5
Age 45–64	64.5	77.3	79.9
Age 65+	69.7	87.1	90.0

Hospital discharges per 1,000

	1964	*1990*	*1995*
Total	109.1	91.0	86.2
Age 45–64	146.2	135.7	122.4
Age 65+	190.0	248.8	266.9

Table 20.1 *continued*

B. Income Comparisons

Interval since last physician contact

	Percent Less Than One Year		
	1964	*1990*	*1995*
Total	66.9	78.2	79.5
Income < $15,000	58.6	77.3	78.2
Income > $50,000	73.6	81.7	83.5
Hospital discharges per 1,000			
	1964	*1990*	*1995*
Total	109.1	91.0	86.2
Income < $15,000	102.4	142.2	140.7
Income > $50,000	110.7	72.5	61.6

Source: Derived from *Health United States* (1998), Tables 77 and 87.

Another measure of elderly access involves hospital discharges per 1,000. In 1964, those 65 and older had 190.0 discharges compared with 146.2 for those aged 45 to 64, a 30 percent differential. By 1990, the differential had grown to 83.3 percent, and by 1995 to 118.1 percent.

Comparing the less affluent to the more affluent in 1964, those with incomes less than $15,000 were 79.6 percent as likely to have seen a physician in the past year as those with incomes higher than $50,000. By 1990, they were 94.6 percent as likely.

In 1964, those with incomes less than $15,000 had 102.4 hospital discharges per 1,000, compared with 110.7 discharges for those with incomes higher than $50,000, or only 92.5 percent as many. By 1990, the lower-income people had 96.1 percent more discharges per 1,000 people, and by 1995 they had 128.4 percent as many.

Health Status

Card, Dobkin, and Maestas (2009) look at the impact of Medicare by examining over 400,000 hospital admissions to the emergency room for "non-deferrable" conditions—diagnoses with the same daily admission rates on weekends and weekdays. There is no discernible rise in the number of admissions at age 65, suggesting that the severity of illness is similar for patients on either side of the Medicare threshold. The insurance characteristics of the two groups differ, however, with a large jump at 65 in the fraction who have Medicare as their primary insurer, and a reduction in the fraction with no coverage. These changes are associated with significant increases in hospital list chargers, in the number of procedures performed in hospital, and in the rate that patients transfer to other care units in the hospital. The authors estimate a nearly 1 percentage point drop in seven-day mortality for patients at

age 65, implying that Medicare eligibility reduces the death rate of this severely ill patient group by 20 percent. The mortality gap persists for at least two years following the initial hospital admission.

Finkelstein and McKnight (2008) remind us that Medicare is a form of insurance against risk. They calculate that the welfare gains from reductions in risk exposure alone may be sufficient to cover between half and three-quarters of the costs of the Medicare program. They view these findings as underscoring the importance of considering the direct insurance benefits from public health insurance programs, in addition to any indirect benefits from an effect on health.

Favorable impacts come with incremental costs. Currie and Gruber (1996) measure the impacts of increased Medicaid eligibility (throughout the United States) for pregnant women between 1979 and 1992. Certain groups saw substantial improvements. For example, a 30 percentage point increase in eligibility among 15- to 44-year-old women was associated with a decrease in infant mortality of 8.5 percent. However, even the most carefully targeted changes in Medicaid eligibility cost the Medicaid program $840,000 per infant life saved, raising important questions of cost-effectiveness. In a similar study, Joyce (1999) finds reductions in newborn costs associated with Medicaid participation (this time in New York) to be between $100 and $300 per recipient, insufficient to offset program expenditures.

Medicare: Recent Changes and Future Prospects

RECENT CHANGES In 1996, trustees of the Hospital Insurance (Part A) Fund predicted that the Part A Trust Fund would have a zero balance by 2001. Uncomfortable with raising payroll taxes, the U.S. Congress chose to make major changes in how Medicare paid health care providers through the Balanced Budget Act (BBA) of 1997.

The BBA increased the incentives for efficient production by mandating the development of prospective reimbursement systems for post-acute care. For hospital outpatient departments, it ended cost-based reimbursement. These two changes virtually ended cost-based reimbursement throughout the Medicare system. Payment formulas for new entrants and for home health services were adjusted downward, and the BBA reduced physician payments.

FUTURE PROSPECTS Despite the major changes in the 1997 BBA it is clear that the U.S. Medicare system will become much larger over the next quarter century. Figure 20.10 displays projections of the Medicare-eligible population, starting in 2015. Projections into the future can be risky, but this one is a safe bet. All those who will be 65 years of age in 2040 are already over 40 years old. To project future populations, demographers statistically "age" the various population cohorts by predicting deaths between now and then. Immigration and emigration generally provide only small adjustments at the national level.

Any way that one looks at things, Medicare will grow. Figure 20.10 shows that the number of Medicare beneficiaries, largely those ages 65 and over, will increase to about 85 million, the result of the baby boom starting in the late 1940s and extending through the early 1960s. In addition, the number of workers (who are also paying into the fund) per beneficiary is projected to fall, from 3.4 in 2010 to 2.3 in 2040. This decrease in workers per beneficiary (occurring in all advanced countries) suggests that there will be more financial pressure on providers to cut costs, and on payers to pay for services.

We must treat projections of future spending with caution, and with the uncertainty surrounding the implementation of ACA, this caution must be re-emphasized. The aging population and expected increases in health care costs suggest a major increase in the Medicare share of the GDP. These projections are based on projected annual growth of GDP of close to 5 percent until 2020, and 4.6 percent thereafter (Medicare Payment Advisory Commission,

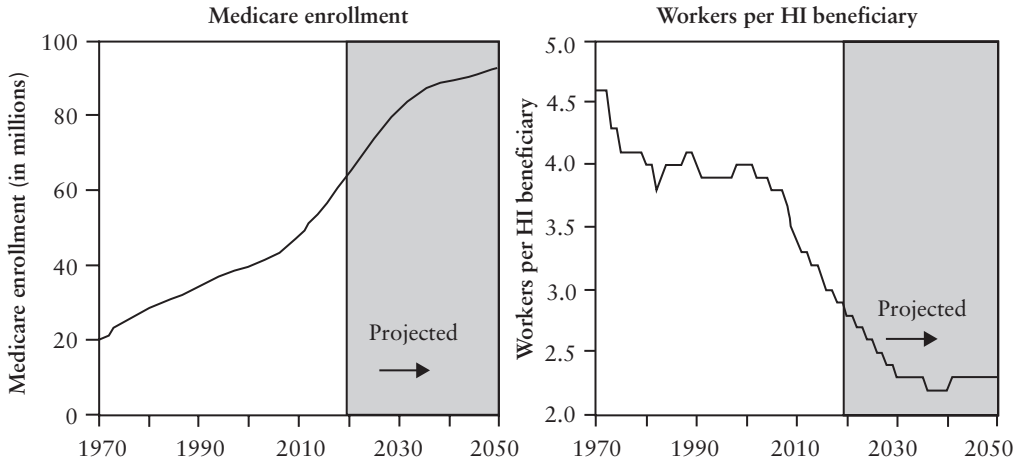


Figure 20.10 Projected Medicare Enrollment and Workers per Enrollee

Note: HI (Hospital Insurance). Hospital Insurance is also known as Medicare Part A.

Source: MEDPAC Report, Chapter 2, June 2015.

2015 Section 2), increasing the denominator of the fraction relating to the percentage of GDP (Percentage = 100 × Expenditures/GDP). Many economists, irrespective of political stripe, would find such projections to be optimistic based on historical perspective. Changing the projected GDP growth rate from 4.6 percent to even a 4.0 percent growth rate, high by historical standards, would raise the 2050 projected share from 5.94 percent to 7.06 percent.

Victor Fuchs (2000) considered three major economic approaches to the crisis implicit in elderly health spending projections. Almost two decades later, his reasoning remains valid.

1 Slow the growth rate of health expenditures. This may be undesirable because many medical advances have improved quality of life for the elderly, and infeasible because the elderly want all of the care that might do them some good.
2 Impose higher taxes on the young to pay for the care for the old. Such tax hikes are not likely, as they will add to an already high burden of support that the young are asked to pay in support of the elderly.
3 Provide more of their own income by increases in work and saving. In earlier work, Fuchs (1999) examined elderly retirement savings, concluding that, "most low-income elderly *could have saved more* [emphasis added] prior to age 65."

Conclusions

In this chapter, we have discussed social insurance and its application to the health care sector. Almost all modern industrialized countries provide fairly comprehensive health care social insurance; we have reviewed the history of these developments and the pattern in the United States.

The major health care social insurance programs in the United States are Medicare, Medicaid, and CHIP. These programs increase health care costs in theory and have been

increasingly costly in practice. It is clear that they have had a beneficial effect on access to care among the elderly and low-income groups, and recent studies suggest that they have a beneficial effect on health status. They leave, however, a substantial number of the poor or uninsured without health care coverage.

The United States has moved toward a national health insurance program with the passage of the Affordable Care Act. Comprehensive social insurance for health care in the United States would directly address and presumably solve the widely perceived problem of providing for the uninsured, a group that often includes people in the poverty, near-poverty, and other lower-income groups. The ACA has provided a partial step in that direction and we discuss it in detail in Chapter 22.

It is also useful to look at health reform beyond the borders of the United States. We begin by comparing the features of health systems across countries in the next chapter.

Summary

1 Several types of social insurance policies and social programs exist, usefully grouped into poverty programs, old-age assistance, disability, health, and unemployment.
2 Social program features include contributions, benefits, length of coverage, means of reimbursement to providers, and methods of determining payment levels to providers.
3 Social insurance originated in nineteenth-century Europe. Social insurance in the United States began with Social Security in 1935 and the adoption of Medicare and Medicaid in 1965. Until 2010, with the passage of the ACA the United States remained one of the few developed countries that had not adopted a comprehensive health care social insurance program.
4 Medicare is a national program that provides hospital insurance to the elderly, along with optional supplemental physician care insurance. The Balanced Budget Act of 1997 established several new categories of Medicare options, and prescription drug coverage (Part D) was passed in 2003 and started in 2006.
5 Medicare Part D, starting in 2006, has led to much increased prescription drug coverage. In 2014, Part D provided $86.4 billion in benefits to 39.2 million enrollees. The average benefit per enrollee exceeded $2,200.
6 Medicaid programs are funded through matching state and federal funds and run by the states. They provide health care to certain categories of the needy and are the primary providers of nursing home aid.
7 Expansion of Medicaid programs serves as one of the two primary enrollment increasing instruments of the Affordable Care Act.
8 Medicare and Medicaid expenditures have increased rapidly since the programs began, due to increased medical care prices, population covered, and quantity of care per capita consumed by the population, as well as due to changes in the nature of the services provided.
9 Medicare and Medicaid accompanied clear improvement in access to care by the lower income population, as evidenced by increased utilization rates by lower-income groups, both absolutely and relatively, to the higher-income groups.
10 With the aging of the baby boom cohort and the improvements in health care technologies in this second decade of the twenty-first century, Medicare must determine how best to structure, provide, and finance the benefits that it is providing to this growing segment of the population.

Discussion Questions

1 In what ways does social insurance differ from private insurance?
2 Of the five types of social insurance programs described, which types characterize Medicare? Which types describe Medicaid?
3 What are the similarities between Medicare and Medicaid? What are the differences?
4 What factors contributed to the historical growth in Medicare spending?
5 Describe how Medicare has affected access to care for the elderly.
6 Does access to health care provided through social insurance programs affect health status? Discuss the evidence.
7 What are some possible reasons that other industrially advanced countries have far more comprehensive social insurance programs for health care than does the United States?
8 Historically in the United States, what groups have supported social insurance for health care, and what groups have opposed it? Why do you think this is the case?
9 For students in the United States, compare Medicaid coverage in your state with coverage afforded to recipients in a neighboring state. Are they the same? If not, why do you think that they differ?
10 The ACA has engendered particular debate between proponents and opponents. Why do you think this is the case?
11 Reductions in federal stimulus plans and decreasing state resources have affected the Medicaid program. Discuss the impact on Medicaid and state responses in the state where you live or go to school.
12 Are elderly people provided the right amount of health care under current U.S. policies? Contrast your answer for the United State to other countries with which you are familiar.

Exercises

1 Calculate the average tax rate for Social Security at incomes of $25,000, $50,000, $75,000, $100,000, $125,000, and $150,000. Do the same for Medicare. (Hint: You may choose to do each graphically.) Characterize each tax as being progressive, regressive, or neutral.
2 Figure 20.3 describes the Medicare Part D prescription drug benefit. Look at the Web page www.partd-medicare.com/ in your area and determine the marginal coinsurance rates applicable in each segment. Then calculate the average amount spent at the following levels of charges: $2,000, $4,000, $6,000, and $8,000. Discuss the "burden" of payments under this schedule.
3 Consider Currie's discussion of take-up of social programs, where the x-axis is program enrollment and the y-axis refers to monetary costs and benefits.
   (a) If we measure the number of people enrolling in a program on the x-axis, why would the "demand" for these programs be downward sloping? Draw a demand curve.
   (b) Why would the costs of establishing a program be upward sloping? Draw a supply curve.
   (c) What is meant by the equilibrium where supply equals demand?
   (d) How can one model program "stigma," and what does it do to equilibrium enrollment? Why?
4 Consider a population of 1,000 families: 200 had Medicaid insurance, 700 had some other type of insurance, and 100 were uninsured. Suppose now that Medicaid broadens

eligibility rules that would allow an additional 100 families to get coverage. After one year, 250 families now have Medicaid, 675 now have some other type of insurance, and 75 are uninsured.
   (a) Calculate the average take-up and crowd-out both in numbers of families and in rates.
   (b) Calculate the marginal take-up rates occurring due to the eligibility change.
   (c) Has insurance coverage for the population increased? Why or why not?
   (d) Has insurance coverage for all families increased? Why or why not?
5 Consider the analysis described in Figure 20.6. Tom and Dick each earn $25,000 per year. Tom has a spouse and two children, and Dick is unmarried. Health insurance and other goods trade off dollar for dollar (there is no tax advantage to health insurance).
   (a) Where would each of the two be located on the budget constraint, and why?
   (b) Which of the two would more likely take up a health insurance program, such as Medicaid?
   (c) How would your answers to the first two parts change if health insurance were subsidized (as it is) relative to all other goods?

Notes

1 From 1937 to 1949, the tax rate was 1 percent of payroll incomes up to $3,000, a maximum tax of $30 per year!
2 The most current information on these programs is: (1) Medicare Program—General Information, www.cms.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html, accessed February 11, 2016; (2) National Medicaid & CHIP Program Information, www.medicaid.gov/medicaid-chip-program-information/program-information/medicaid-and-chip-program-information.html, accessed February 11, 2016.
3 The Affordable Care Act plans to eliminate the doughnut hole by 2020 by reducing expenses for those whose expenditures are in that interval. Until then, Plan D participants will have relatively large coinsurance rates.
4 With the parameters in the example above, the break-even point is about $1,160.
5 These classes were (1) hypertension drugs Angiotensin-Converting Enzyme (ACE) Inhibitors; (2) Alzheimer's disease treatments; (3) Anti-Depressants; (4) Angiotensin Receptor Blockers (ARBs), also used to treat hypertension; (5) Oral Anti-Diabetics; (6) Osteoporosis treatments; (7) Proton Pump Inhibitors (PPIs) for heartburn, gastroesophageal reflux disease (GERD), and ulcers; and (8) cholesterol drugs HMG-CoA Reductase Inhibitors (Statins).
6 Material gathered from http://kff.org/medicaid/report/medicaid-and-long-term-services-and-supports-a-primer/, accessed March 18, 2016.
7 Data are from www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/MMCO_2015_RTC.pdf, accessed March 24, 2016.
8 Tabulations of the unemployed or uninsured often confuse "ever" with "point-in-time." An annual count of children ever-enrolled will *always* exceed the number enrolled at any point-in-time, if new enrollments and departures occur during the year. The greater the number of new enrollments and departures, the greater will be the difference between the point-in-time and annual ever-enrolled counts. Over the first decade of the twenty-first century, over one-third of CHIP enrollees enrolled at any time during the year were not enrolled at the end of the year.

Chapter 21

Comparative Health Care Systems

In this chapter

- Contemporary Health Care Systems
- The United Kingdom—The National Health Service
- China—An Emerging System
- The Canadian Health Care System
- Different Systems: The Public's Evaluation
- Differences in Health Care Spending across Countries
- Conclusions

We now consider the experiences of other countries in providing large-scale health care for two reasons. First, many other countries have constructed programs that predate U.S. programs by generations and provide variations in programs and experiences worth discovering. Second, and more importantly, even with the passage of the Affordable Care Act, the U.S. system has some gaping holes compared with the coverage extended by many other systems. Understanding the approaches used by other countries helps us to assess our own system.

Contemporary Health Care Systems

Many industrialized countries either provide health care directly through the government or provide publicly funded health insurance with comprehensive coverage. Rather than describing details about the health care programs of dozens of countries, we will characterize the basic types of systems employed and develop a few examples in detail.[1]

A Typology of Contemporary Health Care Systems

Böhm and colleagues (2013) classify 30 advanced health care systems from the Organization for Economic Cooperation and Development (OECD) according to three core system dimensions:

- Regulation.
- Financing.
- Service provision.

provided by three categories of actors:

- state-based—typically employees of the state;
- societal—private nonprofit providers, reflecting a societal element, resembling neither for-profit market participants nor part of the state administration;
- private—market-based actors.

Although these three types of systems and three types of actors can produce 27 combinations, the authors argue that only ten are "plausible." As shown in Table 21.1, they group all but one of the countries into one of five plausible combinations for the year 2008. The United States' classification predates the 2010 passage of the Affordable Care Act, which could change its place in future analyses.

In comparing economic data across countries, Table 21.2 shows per capita health expenditures expressed in U.S. dollars in many countries for 2013–2014. Seeking to make these numbers comparable across countries, experts adjust these figures by the purchasing powers of the local currencies (known as *purchasing power parity* or *PPP*). Other columns show each country's health care spending as a percent of GDP for selected years.

The countries vary substantially. No country spends as much as the United States, either in terms of absolute expenditures ($8,713), or percentages (16.4) of the Gross Domestic Product. Contrast the U.S. percentage with our neighbors Canada (10.2 percent) or Mexico (6.2 percent). Several Western European countries (including Austria, Belgium, Denmark, France, and Germany) spend between 10 to 11 percent. The United Kingdom, in contrast, spends only 8.5 percent of its GDP on healthcare.

Table 21.1 Classification of Health System Types

Healthcare system type	*Regulation*	*Financing*	*Provision*	*Cases*	
National Health Service	State	State	State	Denmark	Sweden
				Finland	Portugal
				Iceland	Spain
				Norway	U.K.
National Health Insurance	State	State	Private	Australia	New Zealand
				Canada	Italy
				Ireland	
Social Health Insurance	Societal	Societal	Private	Austria	Luxembourg
				Germany	Switzerland
Private Health System	Private	Private	Private	U.S.A.	
Etatist Social Health Insurance	State	Societal	Private	Belgium	Poland
				Estonia	Slovakia
				France	Israel
				Czech Republic	Japan
				Hungary	Korea
				Netherlands	
Social-based mixed-type	Societal	Societal	State	Slovenia	

Source: Böhm et al. (2013).

Table 21.2 Health System Indicators for OECD Countries

Country	*GDP per capita (US$ PPP)*	*Health spending per capita (US$ PPP)*	*% GDP Spent on Healthcare*	*Life expectancy at birth*	
	2013[b]	*2013*[b]	*2013*[b]	*1970*[a]	*2013*[b]
Australia	44,976	3,866	8.6	70.8	82.2
Austria	45,082	4,553	10.1	70.0	81.2
Belgium	41,573	4,256	10.2	71.1	80.7
Brazil	15,256	1,471	9.6	58.9	75.0
Canada	42,839	4,351	10.2	72.9	81.5
Chile	22,178	1,623	7.3	62.3	78.8
China	11,661	649	5.6	62.9	75.4
Colombia	12,695	864	6.8	–	75.2

continued

Table 21.2 *continued*

Country	*GDP per capita (US$ PPP)*	*Health spending per capita (US$ PPP)*	*% GDP Spent on Healthcare*	*Life expectancy at birth*	
	2013[b]	*2013*[b]	*2013*[b]	*1970*[a]	*2013*[b]
Czech Rep.	28,739	2,040	7.1	69.6	78.3
Denmark	43,782	4,553	10.4	73.3	80.4
Estonia	25,823	1,542	6.0	70.0	77.3
Finland	39,869	3,442	8.6	70.8	81.1
France	37,671	4,124	10.9	72.2	82.3
Germany	43,887	4,819	11.0	70.6	80.9
Greece	25,854	2,366	9.2	73.8	81.4
Hungary	23,336	1,720	7.4	69.2	75.7
Iceland	42,035	3,677	8.7	74.0	82.1
India	4,175	215	5.1	48.8	66.5
Indonesia	10,023	293	2.9	52.4	70.9
Ireland	45,677	3,663	8.0	71.2	81.1
Israel	32,502	2,428	7.5	71.8	82.1
Italy	35,075	3,077	8.8	72.0	82.8
Japan	36,236	3,713	10.2	72.0	83.4
Korea	33,089	2,275	6.9	62.1	81.8
Latvia	22,958	1,055	4.6	69.8	73.9
Lithuania	25,715	1,573	6.1	70.7	73.5
Mexico	16,891	1,049	6.2	60.9	74.6
Netherlands	46,162	5,131	11.1	73.7	81.4
New Zealand	34,899	3,328	9.5	71.5	81.4
Norway	65,640	5,862	8.9	74.4	81.8
Poland	23,985	1,530	6.4	70.0	77.1
Portugal	27,509	2,482	9.0	66.7	80.8
Russian Fed.	25,247	1,653	6.5	68.3	70.7
Slovak Rep.	26,497	2,010	7.6	70.0	76.5
Slovenia	28,859	2,511	8.7	68.7	80.4
South Africa	12,553	1,121	8.9	52.9	56.8
Spain	33,092	2,928	8.8	72.0	83.2
Sweden	44,646	4,904	11.0	74.8	82.0
Switzerland	56,940	6,325	11.1	73.1	82.9
Turkey	18,508	941	5.1	54.2	76.6
United Kingdom	38,255	3,235	8.5	71.9	81.1
United States	53,042	8,713	16.4	70.9	78.8

Notes: [a] 1970 or nearest year; [b] 2013 or nearest year.

Source: *OECD Health Statistics 2015*, http://dx.doi.org/10.1787/health-data-en, accessed November 2016.

Figures like these, as well as concerns about health care access, have led many to question what Americans are getting for their spending. However, high expenditures may have three meanings:

1 High average level of services.
2 High resource costs for services.
3 Inefficient provision of services.

In examining cross-country differences, we note that high levels of services reflect at least the possibility that populations have chosen to spend their incomes in this fashion. We have noted previously that higher income levels lead to higher consumption levels of all normal goods, including health care. Cross-national studies indicate a substantial responsiveness of health care expenditures to increased income (relatively large income elasticity). U.S. expenditure levels reflect in part the higher per capita income level in the United States.

As shown by the comparative data in Table 21.3, the resources available across countries vary widely. Several countries have more practicing physicians per 1,000 than the United States (2.56). Austria has almost twice as many. In contrast, South Korea, Mexico, and several non-OECD countries have fewer. There is a wide range of practicing nurses across the countries, with Iceland, Norway, and Switzerland having the most.

The countries also vary in using technology, measured in units per million people. In 2013, the U.S. had 40.97 computerized tomography (CT) scanners per million. Relative to most countries this is a large number, although Australia has more (55.94). Similarly, the U.S. has 38.05 magnetic resonance imaging machines (MRI), which is larger than most countries other than Japan. Compare the U.S. with the United Kingdom. Roughly speaking, the U.S. has five times as many CTS units per person as the U.K., and over six times as many MRI installations. We will be comparing expenditure patterns and we will discover the U.K. spends a much smaller share of its GDP on health care.

For a better perspective on the relative success in controlling inflation, examine Figure 21.1. The upward trends in expenditure shares continued into the early 1990s for the United States,

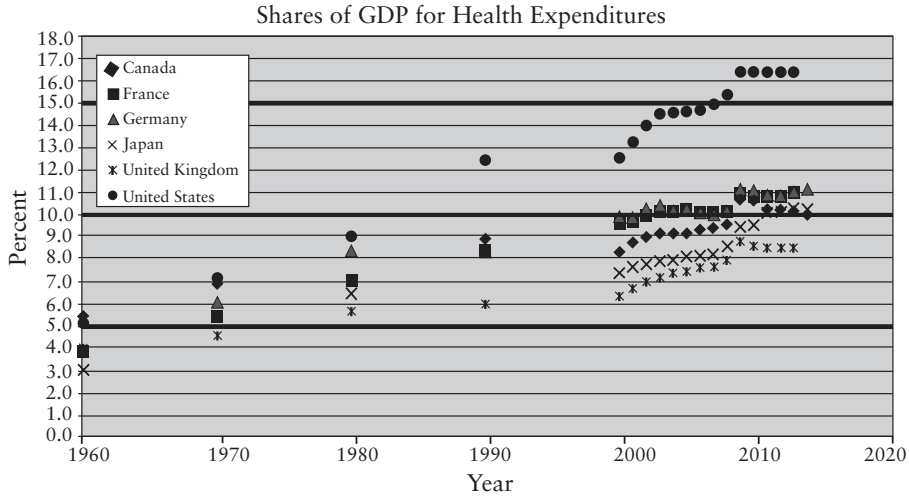


Figure 21.1 Percent of GDP Spent on Health Care, 1960–2014

Note: Pre-2000 percentages are not strictly comparable to percentages 2000 and later.

Source: OECD, 2015.

Table 21.3 Medical Staffing, Equipment, and Technology

2013 or most recent year

	Physician density per 1,000 population 2013	*Nurse density per 1,000 population 2013*	*CT scanners per million population 2013*	*MRI per million population 2013*	*Mammographs per million population 2013*
OECD Countries					
Australia	3.39	11.52	55.94	15.18	22.96
Austria	4.99	7.87	29.60	19.22	23.64
Canada	–	9.48	14.67	8.83	17.33
Chile	–	–	14.76	9.43	12.18
Czech Republic	3.69	7.99	15.03	7.42	11.70
Denmark	–	–	38.06	–	16.62
Estonia	3.28	6.17	18.97	11.38	8.35
Finland	3.02	–	21.70	22.06	28.31
France	3.10	–	15.43	10.94	–
Germany	4.05	12.96	–	–	–
Greece	–	–	35.17	24.30	60.01
Hungary	3.21	6.43	7.88	3.03	14.56
Iceland	3.62	15.45	–	21.83	15.59
Ireland	2.69	–	16.70	13.45	13.23
Israel	3.43	4.87	9.53	2.81	–
Italy	3.90	–	–	24.62	33.47
Japan	–	10.54	–	46.87	31.58

Korea	2.17	5.22	37.09	25.66	54.38
Luxembourg	2.81	11.93	21.58	12.59	8.99
Mexico	2.16	2.62	5.31	2.06	9.24
Netherlands	–	–	11.54	11.49	–
New Zealand	2.81	10.00	17.55	11.18	24.84
Norway	4.31	16.67	–	–	–
Poland	2.24	5.27	17.17	6.44	12.40
Slovak Republic	–	–	15.33	6.65	15.70
Slovenia	2.63	–8.27	13.09	8.74	–16.00
Spain	3.81	5.14	17.59	15.34	15.85
Switzerland	4.04	17.36	36.15	–	–
Turkey	–	–	–	–	11.88
United Kingdom	2.77	8.18	8.10	6.16	8.55
United States	2.56	–	40.97	38.05	–
Non-OECD Countries					
China (PRC)	1.65	2.01	–	–	–
Colombia	1.77	1.03	–	–	–
India	0.73	1.25	–	–	–
Indonesia	0.31	1.15	–	–	–
Latvia	3.19	4.88	34.78	10.43	23.35
Lithuania	4.28	7.55	23.67	10.48	12.65
Russia	4.90	7.43	11.28	3.99	–
S. Africa	0.76	1.21	–	–	–

Source: OECD Health Statistics 2015, http://dx.doi.org/10.1787/health-data-en, accessed November 2016.

and eased some through the 1990s. U.S. expenditures accelerated in the first years of the twenty-first century and jumped in the "Great Recession" of 2008–2009, as did Canada, France, Germany, and the United Kingdom. The recent jump stems in part from a fall in the denominator (GDP per capita) for these countries. Nonetheless, comparing the United States with these other countries shows an increasing spread in expenditure shares, although all have been rising.

The United Kingdom—The National Health Service

This section examines the national health system of the United Kingdom in detail, and the following section looks at China. After that, we will look at Canada and contrast Canada's plan, a national health insurance system, with that of the United States, a private system.

The National Health Service

Great Britain established its National Health Service (NHS) in 1946, and it provides health care to all British residents. About 80 percent is funded by general taxation, with about 19 percent from national insurance and about 1 percent from user charges. Capital and current budget filter from the national level down to the regional and then to the district level. The plan pays general practitioners on a capitation basis and hospital physicians largely on a salaried basis. In addition to the NHS, there is also a private sector health system. About 11 percent of the population purchases private health insurance.

NHS care is largely free at the point of use to all who are "ordinarily resident" in England, as are nonresidents with a European Health Insurance Card. For other people, such as non-European visitors or illegal immigrants, only treatment in an emergency department and for certain infectious diseases is free.

Not all services are free. English patients pay £8.40 (about $11.09 at the August 2016 exchange rate of $1.32 per £1) for each prescription, but close to 90 percent of prescriptions are exempt from charges, and patients in Scotland, Wales, and Northern Ireland are not charged. As of 2016, patients pay no more than £233.70 (about $308) for each "course of [dental] treatment." This maximum, called Band 3, includes crowns, dentures, and bridges—others treatments are far less. Those receiving means-tested benefits and their adult dependents, children under age 16 (under age 19 if a student), pregnant women, and nursing mothers are exempt from dental and prescription charges.

The general practitioner (GP) serves as the gatekeeper to the health care system. GPs are not government employees, but are self-employed and receive about half their incomes from capitation contracts. GPs typically treat routine conditions and refer patients to hospitals for more specialized care. The referral usually will be to a district hospital. Once at the hospital, the patients are under the care of physicians (consultants) who are allocated staffed beds and junior hospital staff to work under their direction.

Table 21.2 shows the U.K. spending per capita ($3,235) in 2013 as 37.1 percent of the U.S. level ($8,713), and a little more than half when expressed as a ratio of GDP (8.5 percent as opposed to 16.4 percent). How does the United Kingdom keep its health care expenditures this much lower while providing universal access to health care? Patients have relatively easy access to primary and emergency care, but specialty care is rationed through waiting lists and limits on the availability of new technologies. A relatively simple model illustrates this phenomenon.

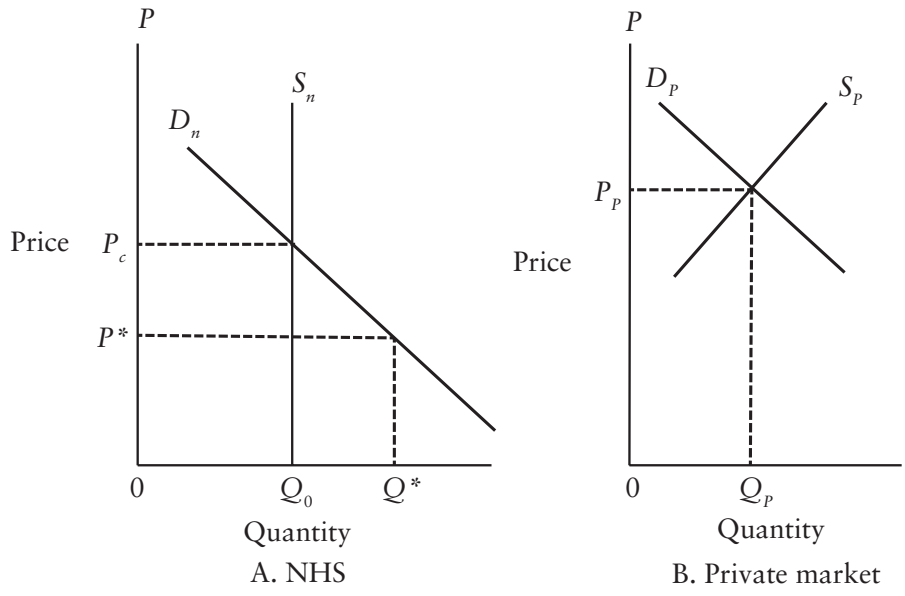


Figure 21.2 Prices and Quantities in a Controlled Market

A Model of Rationed Health Care and Private Markets

We can examine the practices of an NHS-type of organization diagrammatically.

Panel A of Figure 21.2 treats the supply of health services as totally unresponsive to price, or inelastic. Why? The supply curve (a vertical line) indicates that the quantity supplied does not respond to the service price reflecting what the government provides *irrespective* of price. Furthermore, the money price of the services is set by the government at P^*, which is typically less than P_c, the market clearing price. Because of this, we predictably see excess demand ($Q^* - Q_0$) at the administered price P^*. Because most health care cannot be bought and resold (you can't send someone else to get *your* eyes tested or *your* teeth fixed), other rationing, largely waiting time-related, becomes important. For many ailments, NHS patients have faced waiting time (for treatment) periods of months, or even years.

A separate market for services has developed for those who choose to enter the private market without governmental aid, either due to strong preferences for private care or due to the ability to pay more than the NHS price. Returning to Figure 21.2B, excess demand at the administered price P^* represents in part those who are queued and who might wish to pay in the private sector to avoid the long waits. Indeed, some of those in the queue might be willing to pay far more than P^* for the services. Those who participate in the private market, shown in panel B, will pay P_p for the quantity of services, Q_p. The two markets exist simultaneously, although as Box 21.1 indicates, not always comfortably.

PERFORMANCE UNDER THE NHS AND MORE RECENT REFORMS On the one hand, a system such as the NHS that depends on queuing in line for access to care often leads participants to postpone or simply not purchase certain services. On the other hand, the NHS devotes considerable resources to such high-return services as prenatal and infant care (see Box 21.2). To these populations served, and to the larger public seeking equitable provision of care to these segments of the population, the universal nature of the service is particularly beneficial.

BOX 21.1

"Jump the Queue for Cataract Operations by Paying Yourself"

The separation between the English NHS and the private system is not always a large or a comfortable one. In 2015 *Daily Mail* reporter Sophie Borland noted that 41 of 78 England hospital trusts (organizations generally serving either a geographical area or a specialized function such as an ambulance service) offered patients the opportunity to pay for cataract surgery themselves.

More than half of those over age 65 suffered from cataracts, or cloudy patches in the lens that blur or mist the vision. A simple 45-minute operation, with doctors using ultrasound waves to break up the cataract, has dramatically improved sight for millions around the world.

Most hospitals charge patients between £700 and £1,000 per eye for cataract treatment, but the price had risen to £2,552 at Frimley Park Hospital in Surrey and £2,700 at Maidstone and Tunbridge Wells NHS Trust in Kent.

Many NHS trusts impose rules to determine who is eligible for cataract treatment, and routinely turn away patients who cannot read, sew, or watch television. Those patients who do meet the strict criteria must often wait eight months for treatment, over which time their eyesight may deteriorate further and impair basic tasks and hobbies.

The providers actively seek private-pay clients. The University Hospital Southampton's website informs patients that "surgery will be offered much sooner than the usual NHS wait." It adds: "Our cataract choice service offers a new option, between the traditional private sector and the NHS, bringing private healthcare within the reach of many more people." North Cumbria offers patients free parking and a daily newspaper, while at Frimley Park they can choose meals from an "exclusive a la carte menu."

Source: Borland, Sophie "Jump the Queue for Cataract Operations by Paying Yourself: Half of Hospitals Allow Patients to Contribute Themselves (but You'll Pay THREE Times over the Odds)," Sophie Borland, *The Daily Mail*, April 10, 2015, www.dailymail.co.uk/news/article-3032835/Jump-queue-cataract-operations-paying-Half-hospitals-allow-patients-contribute-ll-pay-THREE-times-odds.html, accessed May 13, 2016.

BOX 21.2

How Your Health Visitor Can Help

Many Americans would be surprised to know about the "health visitor" program for newborns and babies in the United Kingdom. Most analysts view this type of program as very beneficial by *marginal benefit–marginal cost* criteria. The following quote describes the services provided.

> A health visitor will usually visit you at home for the first time around 10 days after your baby is born. Until then you'll be under the care of your local midwives.
>
> A health visitor is a qualified nurse who has had extra training. They're there to help you, your family and your new baby stay healthy.

Your health visitor can visit you at home, or you can see them at your child health clinic, GP surgery or health centre, depending on where they're based. They will make sure you've got their phone number.

If you're *bringing up a child on your own* or struggling for any reason, your health visitor can offer you extra support.

Talk to your health visitor if you feel anxious, depressed or worried. They can give you advice and suggest where to find help. They may also be able to put you in touch with groups where you can meet other mothers.

Source: www.nhs.uk/Conditions/pregnancy-and-baby/Pages/services-support-for-parents.aspx, accessed June 16, 2016.

In addition, although the United Kingdom has spent considerably less on health care than the United States and many other countries, by most measures of mortality and morbidity the United Kingdom does about as well. Many nonmedical factors are involved in determining disease and death rates in a population and these factors will vary across countries.

Since 2000, the NHS has faced two major problems. The first relates to capacity constraints—shortages of doctors and nurses, as well as relatively small levels of acute hospital beds. Also, a shortage of nursing home beds has meant difficulties discharging elderly patients from the hospital, preventing hospitals from taking on new admissions. Second, incentive problems pervaded the system. NHS providers were paid salaries to work 11 sessions per week in the NHS. If NHS providers were willing to work (and be paid) for only 10 of the 11 sessions, they were allowed to work as much as they liked in the private sector—where they were paid on a fee-for-service basis. The longest NHS waiting lists occurred in specialties with the highest private earnings.

The NHS has made major efforts to reduce patient waiting times. In 2005, the Healthcare Commission reported that the number of people waiting more than six months for admission as inpatients in England decreased by 85 percent from March 2000 to March 2005. There was also a significant drop in the number of people waiting more than 13 weeks for an appointment as outpatients—down by 92 percent over the same period. For specialties with high inpatient death rates, the number waiting less than six months increased by 8 percent between 1999 and 2005.

However, by March 2007, one in eight NHS hospital patients still had to wait more than a year for treatment. A Department of Health analysis of 208,000 people admitted to the hospital in March showed that 30 percent waited more than 30 weeks and 12.4 percent more than a year. Many people also experienced problems gaining access to NHS dentists, with nearly two-thirds of all dental practices not taking on new NHS patients.

Waiting time problems continue. As of 2016, the NHS actively seeks to limit waiting times to 18 weeks. A 2016 visit to the NHS website shows:

> You have the legal right to start your non-emergency NHS consultant-led treatment within a maximum of 18 weeks from referral, unless you choose to wait longer or it is clinically appropriate that you wait longer.

Moreover:

> Patients with urgent conditions such as cancer and heart disease will be able to see a specialist more quickly. For example, you have the right to be seen by a specialist within a maximum of two weeks *from GP referral* for urgent referrals where cancer is suspected.

The target remains elusive. In April 2015 Haroon Siddique (2015) reported for the *Guardian* that the number and proportion of NHS hospital patients in England waiting more than 18 weeks to begin treatment had risen to their highest levels in almost seven years. In February, nearly 40,000 admitted patients had not started consultant-led treatment within 18 weeks of referral, and more than 13,000 waited more than 26 weeks. Authorities sought for 90 percent to receive treatment within 18 weeks but only 87 percent did so.

The NHS's experience in the area of cost containment seems clear. Rationed care cuts money costs. Even with increased expenditures from the health care reforms, analysts expect total U.K. expenditures to remain well below the European Union and the United States.

China—An Emerging System

The Chinese health economy has undergone substantial changes since the 1949 formation of the People's Republic. Governmental policies moved from a doctrinaire political system with administered prices in the first three decades, to more market-oriented processes since the 1980s, affecting coverage and focus. We begin by describing the Chinese health care economy. We then examine the organization of health care, and the role of the private sector. We then consider some government policy initiatives and measures of system performance. We finish with observations about the future of the Chinese health economy.

China is a large world economy, but its per capita income is small compared to many of the Western countries. Table 21.4 compares the Chinese health economy to developing nations India and Indonesia, as well as to more advanced Japan. According to the World Health Organization (WHO), China spends considerably more on health per capita than do India and Indonesia, but less than Japan. Measures of life expectancy at birth and probability of dying are more favorable than India and Indonesia, but less favorable than Japan.

Table 21.4 Comparative Health Services Data: Four Asian Countries, 2013

Categories	*China*	*India*	*Indonesia*	*Japan*
Total population (in thousands)	1,353,337	1,252,140	249,866	127,144
Gross national income per capita (PPP international $)—World Bank	12,132	5,351	9,752	37,600
Life expectancy at birth male/female (years)	74/77	65/68	69/73	80/87
Probability of dying between 15 and 60 years male/female (per 1,000 population)	103/76	239/158	176/121	81/42
Total expenditure on health per capita ($ 2013)	676	215	293	3,741
Total expenditure on health as % of GDP (2009)	5.6	4.0	3.1	10.3

Source: Health and population data, World Health Organization, www.who.int/countries/en/. Income data, http://data.worldbank.org/indicator/NY.GNP.MKTP.PP.CD/countries, accessed March 3, 2016.

Many are not familiar with general Chinese economic and demographic data. Table 21.5 shows how the population has stabilized (the 35-year-old "one-child" policy ended in January 2016), with an increasing median age, and percentage of elderly. It has also become far more urban with over half the population now living in urban areas. The Chinese economy has exploded, with growth rates greater than 7 percent per year (at this rate GDP would double in ten years). China is currently the second largest economy in the world, although its large population leaves it behind many other countries in per capita terms.

Meng and colleagues (2015) note that since the 1990s, the most significant changes in causes of death in China have been the continuous increase in malignant tumors, cerebrovascular diseases, and heart disease, compared with communicable diseases, chronic respiratory diseases, and digestive diseases (Table 21.6). Many of these are "lifestyle" diseases related in particular to smoking (over 350 million smokers) and dietary considerations. From 1990 to 2010, chronic respiratory tract diseases dropped from top to fourth as a cause of death, with a decline in proportion of deaths from 24.9 percent to 13.5 percent; malignant tumors became the leading cause of death, with their proportion increasing from 19 percent to 26.5 percent.

Describing the delivery system, Eggleston and colleagues (2008a) point to separate urban and rural systems starting in the early 1950s. In urban areas, the three-tier network was composed of street clinics, district hospitals, and city hospitals. In rural areas it consisted of village clinics, township health centers (THCs), and county hospitals. Provincial and central

Table 21.5 Demographic Structure and Social Demographics of China, 1980–2012

Indicators	*1980*	*1985*	*1990*	*1995*	*2000*	*2005*	*2012*
Total population (million)	981.2	1,051.0	1,135.2	1,204.9	1,262.6	1,303.7	1,350.7
Female (%)	48.4	48.4	48.4	48.4	48.3	48.3	48.2
By age:							
0–14 (%)	35.4	30.9	29.3	28.5	25.6	20.5	18.0
15–64 (%)	59.5	63.5	64.9	65.3	67.5	71.8	73.3
Over 65 (%)	5.1	5.6	5.8	6.2	6.9	7.7	8.7
Annual population growth rate (%)	1.3	1.4	1.5	1.1	0.8	0.6	0.5
Population density (number of people/km^2)	105.2	112.7	121.7	129.2	135.4	139.8	144.8
Total fertility rate (%)	2.7	2.8	2.5	1.7	1.5	1.6	1.7
Crude birth rate (per 1,000)	18.2	21.0	21.1	17.1	14.0	12.4	12.1
Proportion of urban population (%)	19.4	22.9	26.4	31.0	35.9	42.5	51.8
GDP (PPP) $ billion	–	–	1,110.0	2,151.4	3,616.3	6,470.2	14,782.7
GDP (PPP) $ per capita	–	–	1,006.6	1,785.6	2,864.1	4,162.9	10,944.5

Source: World Bank, World Development Indicators, 2014.

Table 21.6 The Top Six Causes of Death in China in Selected Years

Rank	*1990*		*1995*		*2000*		*2005*		*2010*	
	Cause	*Pct*	*Cause*	*Pct*	*Cause*	*Pct*	*Cause*	*Pct*	*Cause*	*Pct*
1	Chronic respiratory tract diseases	24.9	Chronic respiratory tract diseases	25.3	*Malignant tumors*	22.3	*Malignant tumors*	26.3	*Malignant tumors*	26.5
2	Cerebrovascular diseases	19.0	*Malignant tumors*	20	Cerebrovascular diseases	21.4	Cerebrovascular diseases	21.7	Cerebrovascular diseases	23.4
3	*Malignant tumors*	19.0	Cerebrovascular diseases	19.7	Chronic respiratory tract diseases	21.4	Chronic respiratory tract diseases	19.5	Heart disease	20.8
4	Heart disease	13.4	Heart disease	12.1	Heart disease	15.3	Heart disease	15.1	Chronic respiratory tract diseases	13.5
5	Injuries and poisoning	8.8	Injuries and poisoning	10.9	Injuries and poisoning	9.7	Injuries and poisoning	8.8	Injuries and poisoning	7.9
6	Digestive diseases	5.2	Digestive diseases	4.7	Digestive diseases	3.9	Digestive diseases	3.4	Digestive diseases	2.7

Source: National Health and Family Planning Commission, 2014.

hospitals provided high-level referral care. Under this system, the Ministry of Health or the local Bureau of Health managed the majority of the provider organizations.

Maoist Communist rule through the 1970s sought to assure access to care. Providers received direct budgetary support to cover the difference between costs and revenues earned from the nominal fees that were paid. The government financed preventive and other public health services and provided anti-epidemic stations at province, prefecture, and county/district levels, as well as at THCs and village clinics.

When the government routinely subsidized the providers, if prices differed from costs, this had little impact because the government could eliminate deficits through subsidies. Without government subsidies, however, in subsequent years, providers have tended to favor high-technology diagnostics at the expense of less-profitable basic services.

The current Chinese health system has systems for

- health financing,
- health service delivery, and
- health supervision.

By 2013, the three insurance schemes covered more than 95 percent of the total population, although benefits vary by insurance scheme due to differing funding levels. The delivery of health care services had previously relied on a system of predominantly public hospitals and other public health care facilities including traditional medicine hospitals. Meanwhile, the growing role of the private sector as supplementary of the public financing and delivery system has been emphasized in health care financing and delivery. Private health insurance expenditure reached 3.6 percent of total expenditure on health in 2013, while private health institutions accounted for 45.1 percent of health institutions.

China (notes Eggleston, 2012), as with other East Asian countries such as Japan, South Korea, or Taiwan, has never had gatekeeper requirements. Patients can traditionally self-refer to any provider, although social health insurance programs limit coverage for providers outside the given locality (county or municipality).

The Chinese system has featured a continuing disparity between urban and rural care. The adoption of more market-based policies led to wide divergence in health-related resources. In 2012, write Meng and colleagues (2015), there were 8.54 health care professionals per 1,000 population in urban areas and 3.41 in rural areas. The number of beds per 1,000 people (2012) is 6.88 in urban areas, compared to 3.11 in rural areas.

The 1990s saw the initiation of several new policies in both urban and rural areas. In urban areas, municipal risk pooling for employees, known as Basic Medical Insurance (or BMI), was established. The government also established a series of medical savings accounts, but they did not stipulate the means of provider payment. As a result, most people purchase treatment under a fee-for-service (FFS) model. In rural areas, the government established a new cooperative medical scheme (NCMS), which combines household contributions with central and local government subsidies. It was piloted in 2003, and Chen and colleagues (2011) report that 95 percent of the counties were implementing the scheme by 2008.

In the 1980s and 1990s, with an absence of universal health insurance coverage and the low coverage of basic medical insurance, health expenditure largely consisted of out-of-pocket (OOP) payments. Recent government investment in health and establishment of basic medical insurance have reduced OOP health payments and raised the accessibility and equity of health services.

The Chinese have reacted strongly to public health emergencies, particularly the severe acute respiratory syndrome (SARS) epidemic in 2003. Decision makers reacted to perceived

inadequate public health inputs, poor rural health infrastructure, and weak risk-pooling capacity of medical security schemes.

Basic health insurance schemes saw public funding increased threefold to reach 320 yuan (about $48) per capita in 2014. The share of public funds as a percentage of the total expenditure on health reached 30 percent. China has also sought to allow private health care facilities to enter more areas of service provision and improve the policy environment where these facilities operate. Between 2009 and 2012, the number of private hospitals increased from 6,240 to 9,786, and the overall number of nonpublic hospitals increased from 10,640 to 13,533. Meng and colleagues (2015) note that compared with the second stage of health reform (1979–2002), government policy post-2003 has encouraged capital to enter the health market, and has given support for the nonprofit focus of public health facilities and for preventive care.

All of these developments have led to a rapidly growing health care system. Between 1995 and 2012, China's total health expenditure increased by a factor of 12.9, with the percentage of GDP growing from 3.5 percent to 5.4 percent. In 2012 for example, 30.0 percent of the total health expenditure was covered by the government budget, 35.6 percent by social insurance, and 34.4 percent by OOP payments. Eggleston (2012) summarizes coverage of this early twenty-first-century Chinese health economy as "wide" and "shallow."

The Canadian Health Care System

Rapid increases in U.S. health care costs and pre-ACA concern over the large number of uninsured led many in the United States to look at Canada's health system as a model for reform. Many Americans perceive that Canada has developed a comprehensive and universal national health insurance program that is cost-effective and highly popular.

Background

Canada and the United States share a long border and similar heritage in terms of language, culture, and economic institutions.[2] The health care systems evolved similarly until the 1960s and as recently as 1971, both countries spent approximately 7.5 percent of their GDPs on health care.

After 1971, however, the health care systems moved in different directions. While Canada has had publicly funded national health insurance, the United States has relied largely on private financing and delivery (although government has been heavily involved through Medicare, Medicaid, and numerous regulatory programs). During this period, spending in the United States grew much more rapidly despite large groups that were either uninsured or minimally insured.

The Canadian system of financing and delivering health care is known as Medicare, not to be confused with the U.S. Medicare program for the elderly. In Canada, each of the ten provinces and three territories administers a comprehensive and universal program partially supported by grants from the federal government.

The Canada Health Act of 1984 defines five principal features to all Canadian health care. Each provincial health care insurance plan must be: (1) publicly administered; (2) comprehensive in coverage; (3) universal; (4) portable across provinces and outside the country; and (5) accessible (i.e., without user fees and with free choice in the selection of providers).

Allin and Rudoler report that in 2014 about half of all practicing physicians (2.24 per 1,000 population) were general practitioners, or GPs (1.14 per 1,000 population), and half (1.10 per 1,000 population) were specialists (Canadian Institute for Health Information, 2015b). Primary care physicians generally serve as gatekeepers, and many provinces pay lower fees to specialists for nonreferred consultations. Most physicians are self-employed in private practices and paid fee-for-service, although there has been recent movement toward group practice and alternative forms of payment, such as capitation (per person, rather than per service) models.

Canadian hospitals are private and generally not-for-profit institutions, although their budgets are approved and largely funded by the provinces. However, Allin and Rudoler report that some provinces have introduced activity-based funding to pay for additional services targeted to reduce waiting times. Ontario, for example, adopted activity-based funding for cataract, joint replacement, and cardiac bypass surgery, and has successfully reduced waiting times.

Two key provisions of the 1984 Canada Health Act guide Canada's Medicare:

- no extra billing by medical practitioners or dentists for insured health services under the terms of the health care insurance plan;
- no user charges for insured health services by hospitals or other providers under the terms of the health care insurance plan.

The provinces and territories also provide coverage to certain groups of people (e.g., seniors, children, and social assistance recipients) for health services that the publicly funded health care system does not generally cover. These supplementary health benefits often include prescription drugs, vision care, medical equipment and appliances (prostheses, wheelchairs, etc.), independent living, and the services of podiatrists and chiropractors. Dental services are much like services in the United States—either uninsured, privately insured, or group insured through place of employment. The level of health coverage varies across the country. Many Canadians have supplemental private insurance coverage, through group plans, which covers the cost of these supplementary services.

Table 21.7 provides comparative data on the two countries. While geographically larger than the United States, Canada has about 11 percent of the U.S. population. Canada's GDP per capita is about 80.8 percent of the U.S. level. With a national health system providing universal coverage, public funds account for over 67 percent of total health spending. Canada has maintained substantially lower health spending and share of GDP per capita than the United States, despite its universal health insurance system and its longer lengths of stay.

According to the World Bank, Canada has about 2.1 practicing physicians per 1,000 people, compared with 2.4 in the United States, and about 9.5 nurses per 1,000 compared to 9.8 in the United States. A considerable portion of the U.S. population has gone without insurance coverage, even with the recent passage of the Affordable Care Act, but Americans spend almost twice as much per capita on health care ($9,086 versus $4,569). Canadians drink a little less alcohol and smoke a little more than do Americans. Despite lower spending, Table 21.7 shows that life expectancy, a commonly compared health status indicator, is about 4.3 years longer for women and 4.8 years longer for men in Canada. Finally, public opinion polls indicate that Canadians support their system more than Americans support theirs and are concerned about any threats to it. Given the Canadian record on cost savings, health care scholars, policymakers, and politicians have shown great interest in determining the sources for its apparent success.

Table 21.7 Comparative Data: Canada and the United States, 2013–2015

	Canada	*United States*
Population—2013 in millions[a]	35.3	316.1
Population over 65 (2013, %)[a]	15.2	14.1
GDP—2010 (trillions of 2014 $US)[b]	1.57	17.42
GDP per capita—2014 (2014 $US)[b]	44,100	54,600
Health spending per capita—2013 ($US PPP)[a]	4,569	9,086
Health spending—2013 (% of GDP)[a]	10.7	17.1
Percent of total health spending (2013[a])		
Public expenditures	67.3	46.2
Inpatient care	22.0	18.0
Outpatient care	34.0	52.0
Pharmaceuticals	16.8	11.7
Acute care inpatient beds/1,000 population (2013[a])	1.7	2.5
Average length of stay (acute care days) (2013[c])	7.6	5.4
Uninsured population in percent (2015[c])	0.0	11.4
Out-of-pocket payments per capita ($US)—2013[a]	623	1,074
Tobacco (% population 15+)—2013[a]	14.9	13.7
Alcohol consumption (liters/capita 15+)—2012[a]	8.1	8.8
Life expectancy (in years) at birth—females (2011[a])	83.6	81.1
Life expectancy (in years) at birth—males (2011[a])	79.3	76.3

Source: [a] OECD *Health Data 2015*, March 2016; [b] World Bank, http://data.worldbank.org/indicator/, 2015, accessed November 2016; [c] Gallup poll www.gallup.com/poll/184064/uninsured-rate-second-quarter.aspx, accessed November 2016.

Physician Fees and Quantity

Table 21.7 shows that Americans spend about twice as much on health care as do Canadians. Do Americans get twice the level of services (provider visits, hospital days, pharmaceuticals)? Do providers charge twice as much for services? Are the services twice as good in the U.S.? Because health care is a complicated bundle of literally thousands of potential components, these questions have required careful analysis of these component parts.

Fuchs and Hahn (1990) sought to break down Canadian and U.S. expenditures by specific services. The authors estimated that 1985 spending on health care per capita was 38 percent higher in the United States. More striking was the disparity in spending on physician services: 72 percent higher in the United States, and 178 percent higher for the procedures component.

With aggregate spending equal to the product of prices and quantities, the authors wanted to identify differences in fees (prices) and utilization per capita (quantities). Overall, fees were 239 percent higher in the United States for 1985. Though there were variations in the ratios across service categories, U.S. fees were considerably higher in each category. The net incomes of U.S. doctors were also substantially higher than were their Canadian counterparts.

To explain the smaller Canadian fees, we observe that Canadian provincial governments constitute monopsonies (single buyers) of physician labor, and some feel that this reduces overall spending. Monopsony means that the provincial governments face upward-sloping supply curves for physicians, so that the marginal labor cost of raising the fees for one physician requires raising the fees for all others. This results in lower fees than with competitive buyers, and in hiring fewer workers than in a competitive market (readers can look ahead to Figure 21.3C for a monopsony analysis). Negotiations with the local medical societies reflect this monopsony power as compared to the United States with its myriad buyers.

The differences in service volume found by Fuchs and Hahn were perhaps more surprising than the fee differentials. Despite the much higher spending per capita for physician care, the quantity of care per capita was considerably lower in the United States. Thus, the savings in Canada, at least for physician care, did not come from reduced volume of care.

The Fuchs and Hahn findings provided provocative insights and led readers to numerous questions about the two systems. After discussing the Canadian system in more detail, we return with a study by June and Dave O'Neill (2008) that revisits some of the questions.

Why Are Fees and Hospital Costs Lower in Canada?

Hospital patients in Canada have longer lengths of stay, in part because of the greater use of Canadian hospitals for chronic long-term care. Nonetheless, after adjusting for differences in case mix between the two countries, Newhouse, Anderson, and Roos (1988) found that the cost per case-mix adjusted unit was roughly 50 percent higher in the United States. Several reasons may explain this phenomenon.

In Canada, unlike the United States, physician fees result from negotiation between physicians' organizations and the provincial governments, as well as from other limits on total spending. Physicians cannot evade the fee controls by charging extra (sometimes called *balance billing*) to patients who can afford it.

The provinces also regulate hospital costs similarly through approval of hospital budgets. Hospitals and provinces negotiate operating budgets financed by the provincial governments. The capital budget may include other sources of funding, but provinces still must approve capital expenditures. Thus, a centralized mechanism allocates resources to the hospital sector and determines the distribution of resources among hospitals. Occupancy rates are higher in Canadian hospitals. Also, returning to Table 21.3, looking at CTS, MRI, and mammographs, provinces have limited the capital costs associated with expensive new technologies.

Administrative Costs

One of Canada's major cost advantages involves administrative and other overhead expenses. Almost all U.S. patients have experienced coverage or billing problems due to extensive and complex paperwork practices. Patients moving among providers must provide the same information to multiple providers multiple times. With inaccurate transmittal of data, inconvenience can turn into something much worse if providers make inaccurate decisions based on incomplete or inaccurate data. For providers and third-party payers, too, the paperwork is not simply inconvenient, but expensive, as it involves major personnel and data systems allocations.

Woolhandler, Campbell, and Himmelstein (2003) compared 1999 Canadian and United States administrative costs and calculated U.S. excess per capita administrative costs of $752, or $209 billion in aggregate. They argued that a single-payer, Canadian-style health system for the United States would save $0.71 out of every $1 of U.S. administrative costs.

Re-examining their data, Aaron (2003) argued that analyzing per capita expenditures overstates the difference because it depends on arbitrary assumptions relating to currency values

and wage rates. According to Aaron, administrative costs in the United States accounted for about 31 percent of total health care spending compared to 16.7 percent in Canada. This 14.3 percentage point differential, if applied to the United States, would save 46.1 percent (i.e., 14.3 divided by 31) of U.S. administrative costs (compared to Woolhander et al.'s 71 percent), or $489 per capita. However, even this more conservative calculation pointed to excess spending at that time of $159 billion per year!

Subsequent analyses verify this result. Pozen and Cutler (2010) break down the $1,589 difference in 2002 health expenditures per capita between the U.S. and Canada. Adjusting for population size, there are 44 percent more administrative staff in the U.S. system than in the Canadian system. The authors find that higher administrative costs in 2002 accounted for $616 or 39 percent of the difference. A 2012 calculation by the textbook authors multiplied that figure by 310 million Americans, yielding a total of $232 billion dollars in "excess" administrative costs, or between 8 and 9 percent of total U.S. health expenditures. A 2014 article by Himmelstein and colleagues validates these potential savings for the U.S. hospital sector, noting that 25 percent of all U.S. hospital spending consists of administrative costs, compared with 12 percent for Canada and Scotland (other nations fall in between).

Administrative costs are real costs, and defenders of the U.S. multiple insurer system might argue that it provides varieties of coverage to match consumer preferences. However, the incrementally large administrative costs of a multiple insurer–multiple payer system in the U.S. system do not appear to bring commensurate benefits.

A Comparison

The foregoing data suggest that the Canadian system is more effective than the U.S. system in several respects. Costs are lower, more services are provided, financial barriers do not exist, and health status as measured by mortality rates is superior. Canadians have longer life expectancies and lower infant mortality rates than do U.S. residents.

However, the Canadian system has had its own financial problems. As a result of unprecedented federal deficits in the 1990s, the Canadian government substantially reduced its cash transfers to the provinces. Despite considerable improvement in the federal government's fiscal health, provinces must find new sources of tax revenue, impose more stringent fee and budgetary controls on health providers, increase efficiency in health care delivery, scale back on benefits by no longer insuring some previously covered services, and impose user fees.

Similar to the United States, the provinces have forced large reductions in hospital capacity with a corresponding substitution of outpatient care for inpatient care. Regional boards with budgetary authority have replaced centralized provincial departments.

Watson and Allin (2016) report that sub-national cost-control measures include mandatory annual global budgets for hospitals and health regions, negotiated fee schedules for health care providers, drug formularies, and reviews of the diffusion of technology. Further, many governments have developed pricing and purchasing strategies to obtain better drug prices. In July 2010, the ten provinces and three territories agreed to establish a "pan-Canadian" public sector purchasing alliance of common drugs and medical equipment and supplies.

System critics charge that health care is rationed in the sense that all the care that patients demand, or would be provided to meet their best interests, cannot be supplied on a timely basis. As noted in the model on the British NHS, rationing below market price leaves some people, who would be willing to pay more, unable to purchase any of the good at all.

Comparative international surveys indicate that capacity limits and new technology result in longer waiting periods (Table 21.8) for elective surgery in Canada. The "safety valve" of a

Table 21.8 Selected Health System Performance Indicators for 11 Countries, 2013 or Most Recent Year

		Australia	*Canada*	*France*	*Germany*	*Netherlands*	*New Zealand*	*Norway*	*Sweden*	*Switzerland*	*U.K.*	*U.S.*
a. Adults' access to care, 2013	Able to get same-day/next-day appointment when sick	58%	41%	57%	76%	63%	72%	52%	58%	n/a	52%	48%
	Very/somewhat easy getting care after hours	46%	38%	36%	56%	56%	54%	58%	35%	49%	69%	39%
	Waited 2 months or more for specialist appointment	18%	29%	18%	10%	3%	19%	26%	17%	3%	7%	6%
	Waited 4 months or more for elective surgery	10%	18%	4%	3%	1%	15%	22%	6%	4%	n/a	7%
	Experienced access barrier because of cost in past year	16%	13%	18%	15%	22%	21%	10%	6%	13%	4%	37%
b. Avoidable deaths, 2013	Mortality amenable to health care (deaths per 100,000 population)	68[a]	78[a]	64[a]	88	72	89[a]	69	72	n/a	86	115[b]

continued

Table 21.8 *continued*

		Australia	*Canada*	*France*	*Germany*	*Netherlands*	*New Zealand*	*Norway*	*Sweden*	*Switzerland*	*U.K.*	*U.S.*
c. Public views of health system, 2013	Works well, minor changes needed	48%	42%	40%	42%	51%	47%	46%	44%	54%	63%	25%
	Fundamental changes needed	43%	50%	49%	48%	44%	45%	42%	46%	40%	33%	48%
	Needs to be completely rebuilt	9%	8%	11%	10%	5%	8%	12%	10%	7%	4%	27%

Notes: [a] 2011; [b] 2010.

Source: Elias Mossialos, Martin Wenzl, Robin Osborn, and Dana Sarnak, *2015 International Profiles of Health Care Systems*, Commonwealth Foundation (2016), www.commonwealthfund.org/publications/fund-reports/2016/jan/international-profiles-2015, accessed June 7, 2016.

private system, like the United Kingdom, for those willing to pay more is not readily available, although some Canadians (particularly those near large U.S. border cities such as Buffalo and Detroit, or with winter homes in Florida) use U.S. facilities for this purpose.

Solid research comparing the two systems has been sparse. This section presents two papers that attempt to decompose the differences. June and Dave O'Neill (2008) raise three questions regarding American and Canadian differences:

1 What differences in health status can be attributed to the two systems?
2 How does access to needed health care resources compare?
3 Is inequality in access to resources different?

They use a data set from the Joint Canada/U.S. Survey of Health (JCUSH), designed and conducted jointly by Statistics Canada and the U.S. National Center for Health Statistics, which asked the same questions to representative samples of U.S. and Canadian residents.

The authors examine various aggregate data sources to break down differences in life expectancy (Canadians live longer) and causes of mortality (Americans are more susceptible to nondisease determinants such as accident and homicide). Americans tend to have lower birthweight babies (with higher mortality rates), and Americans tend to be more obese. The authors argue that while health care systems (including better prenatal care) can help, they have much less direct effects on important factors such as obesity.

Regarding unmet needs, the authors use the JCUSH to examine difficulty in receiving health care. Those with an unmet need were asked the reason for the unmet need—had to wait too long or service not available; cost (i.e., could not afford service); or a reason other than those two. The "wait too long/service not available" reason (56.3 percent) dominated among the Canadians who had an unmet need, while for U.S. residents cost (54.7 percent) was the major factor and "waiting too long" (13.2 percent) was relatively minor.

The authors also examined the effect of unmet needs on the Health Utility Index (HUI) of health status and found that in Canada unmet needs reduced the HUI by 0.097 (compared to a mean of 0.898) when the individual cited waiting as a reason for unmet need. The effect was much smaller and not significant for Americans.

Inequality in access relates people's scores on the HUI to income. The HUI provides a description of overall functional health based on eight attributes—vision, hearing, speech, mobility (ability to get around), dexterity of hands and fingers, memory and thinking, emotion, and pain and discomfort. If a single-payer system equalized health irrespective of income, one would expect income to show a zero impact. Comparisons of subjects in the U.S. and Canada showed the relationship of health to income to be *roughly similar* in the two countries.

The authors conclude that the U.S. and the Canadian systems provide similar results. The need to ration "free" care may ultimately lead to long waits or unavailable services and to unmet needs. In the United States, costs often lead to unmet needs, but costs "may be more easily overcome than the absence of services." When those aged 18 to 64 were asked about satisfaction with health services and the ranking of the quality of services recently received, more U.S. residents than Canadians responded that they were fully satisfied (51.5 percent vs. 41.3 percent) and ranked quality of care as excellent (40.4 percent vs. 37.7 percent). Satisfaction and quality of care may relate to expectations as well as to objective measures.

The authors do not address the differential in per capita health care expenditures, which at the time were over 80 percent higher in the United States. They ask, "Is the U.S. getting sufficient additional benefits to justify these greater expenditures and where should we cut back if cutbacks must be made? Alternatively, what would Canada have to spend to increase their

technical capital and specialized medical personnel to match American levels or to eliminate the longer waiting times? And would it be worthwhile to them to do so?"

Two Canadian scholars, Duclos and Échevin (2011), address the O'Neill income-health relationship in more detail. Using alternative analytical methods ("stochastic dominance") they rank Canada and the U.S. using data from the Joint Canada/United States Survey of Health. They find that Canada dominates the United States over the two groups of lower health statuses in terms of the bi-dimensional distribution of health and income. This occurs because Canada has better health distribution, a lower correlation between income and health, and lower income inequality.

Different Systems: The Public's Evaluation

Decisions about health care systems ultimately reflect the attitudes of the public with regard to satisfaction, cost of care, and quality of care. Schoen and colleagues (2016) surveyed citizens of Australia, Canada, France, Germany, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States who had had recent experience with their countries' health care systems, regarding general satisfaction, access to care, cost of care, and quality of care. All of the countries are economically advanced, but they have a wide range of insurance and care systems. The surveys used common questionnaires translated and adjusted for country-specific wording.

Although the researchers evaluate many health economy dimensions, we focus on issues of access, avoidable deaths, and satisfaction. As we noted earlier, satisfaction with a system comes both from expectations and system performance. Different people may register different levels of satisfaction with the same services and outcomes, depending on their expectations.

Table 21.8 shows that access (item a) varies according to the service needed. Participants in the German system reported a high degree of same-day/next-day appointment success (76 percent), whereas Canadians were a little more than half as likely (41 percent) to see a provider either the same or the next day, with the United States the next lowest at 48 percent. In contrast, several countries had very small numbers (France, Germany, the Netherlands, Switzerland, and the United States in single digits) waiting four months or more for surgery. Canadians (18 percent) and Norwegians (22 percent) had the highest rates. Box 21.3 provides a Canadian example.

BOX 21.3

"Someone Else Needed It Before I Did"

Expectations are critical in evaluating health system outcomes. While vacationing in Florida, one of the authors (Goodman) played golf with a Canadian who remarked that this was his first round after having had his hip replaced. The surgery had incurred no out-of-pocket costs, and he felt fine. When asked how long he had to wait for surgery, he responded "18 months." Did the wait bother him? "No . . . it was free when I got it, and someone else needed it before I did."

Avoidable deaths per 100,000 people (item b) refer to the system's ability to respond to health care needs. While it is impossible to bring the level down to 0, a well-functioning health care system would reduce it. Of the 11 countries surveyed, avoidable deaths vary from 64 for France to 115 for the United States.

Item (c) of Table 21.8 suggests that in most countries close to half of the population view their system as working "well," with U.K. citizens having the highest percentage—63 percent. Many citizens among the 11 countries view fundamental changes as needed, but most of those surveyed do not believe that the system needs to be rebuilt. Survey data on the need for rebuilding varies from 4 percent for the U.K. to 27 percent for the U.S.

Differences in Health Care Spending across Countries

Different countries have different incentive systems, and, in fact, have differing shares of national product in the health care sector. Having described the systems, and examined the health sector shares of national product, it is appropriate now to explore why the shares differ.

A Model of Health Expenditure Shares

Consider a model of health expenditures and call total expenditures on health care E. By definition, these expenditures equal the price of health care multiplied by the quantity of health care consumed, or $E = PQ$. Defining the share of national income spent on health care as s, we calculate s as the ratio of E to national income, Y, or:

$$s = PQ/Y \tag{21.1}$$

We have seen that share, s, can increase because either the price or quantity has increased, or because the national income has decreased. In fact, mathematically:

$$(\%\ Change\ s) = (\%\ Change\ P) + (\%\ Change\ Q) - (\%\ Change\ Y) \tag{21.2}$$

So, for example if the price of care increases by 1 percent, the quantity decreases by 0.5 percent, and income stays constant, the share will increase by (+0.1 – 0.05), or +0.05 percent.

Although the preceding expression is an identity, mathematically true by definition, it can provide useful insights. If the price of health care, P, increases by the same rate as all other prices, the health care share of national income does not change because percent change in price is offset by percent change in national income (prices multiplied by quantities). If health expenditures pQ increase at the same rate as income y, again health care share does not change.

APPLYING THE MODEL Rather than looking just at the percentage changes that occur, we try to examine why. Suppose the prices of health care relate to the kind of health system the country has or to the social insurance scheme. Also, recognize that the quantity of health care used, Q, tends to increase when national income, Y, increases. Note further that through the demand relationship, quantity of health care, Q, is negatively related to the price of health care, P.

Consider several ideas in turn:

1 An increase in health care price would increase the share if there were *no* consumer response. The extent to which consumers reduce quantity demanded (in response to price changes) will offset the increase in prices.
2 An increase in the share of population who use health care would tend to increase health care expenditures and the share of GDP going to health care.

3 An increase in national income, *Y*, unaccompanied by an increase in health care demand would decrease the share. However, if increased income leads to increased demand, the effect depends on the demand elasticity. A 1 percent increase in national income that leads to a 1 percent increase in expenditures (that is, the income elasticity equals +1.0) will result in a constant share.

Moreover, market structure matters! Economists often implicitly view expenditures in the context of perfectly competitive markets. If valid, as noted in Figure 21.3A, then the total health expenditures box (the numerator of fraction of GDP going to health care) accurately reflects the resource costs P^* of health care at the margin. Anderson and colleagues (2003), however, argue that the markets for the health workforce (especially physicians) are still largely national and even local within countries. Moreover, many markets related to health care within localities do not satisfy the rigorous conditions of the textbook model of competition.

We find varying degrees of monopoly power on the "sell" side of the market and varying degrees of monopsony power on the "buy" side. Because monopolists (Figure 21.3B) equate marginal costs to marginal revenues, they can raise prices above those they would obtain in perfectly competitive markets. This earns them "economic rents," defined as the *excess* of the prices actually received by sellers above the minimum prices the sellers would have to be paid to sell into the market. Figure 21.5B shows that the resource costs (the box defined by the supply curve) are considerably less than the total expenditures (the sum of the resource costs and the monopoly rents), with the difference going as rents to providers. Monopoly quantity Q_b is also less than Q^* under competitive markets because in order to increase prices, monopolistic providers sell less.

Some countries try to reduce the rents earned on the supply side by creating market power on the buy (monopsony) side of the market. A single-payer system (similar to the one used by Canadian provinces) would be related to a "pure monopsony." A pure monopsonist (Figure 21.3C) must pay increased resource costs to all supply factors, so the monopsonist faces a market marginal cost curve, not unlike the monopolist's marginal revenue curve. Here, the producer provides quantity Q_c, but expenditures are much smaller than in Figure 21.3B.

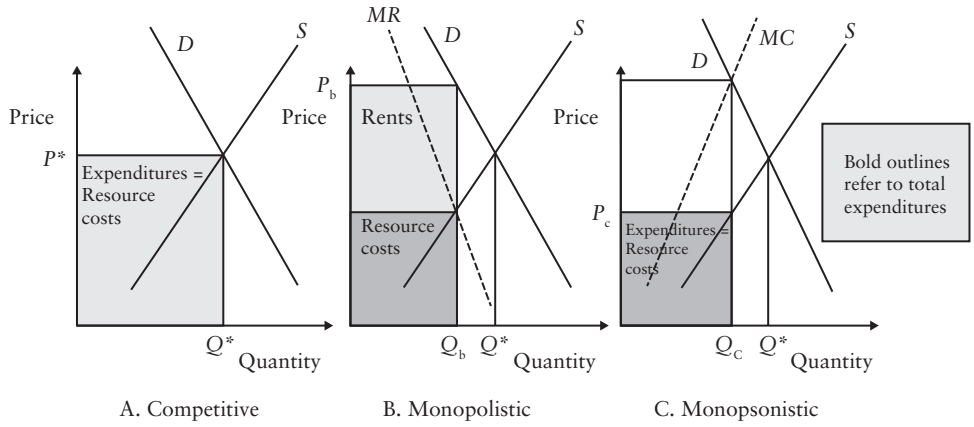


Figure 21.3 Health Expenditures by Market Structure

Note again that in *either* the monopolistic or the monopsonistic case, the quantity of services provided falls short of the optimum Q^*. We have intentionally drawn the Figure 21.3 monopoly and monopsony quantities Q_b and Q_c to be identical, but in the monopolistic case, extra resources are *transferred* as monopoly rents from the buyers to the sellers.

The U.S. Medicare program and Medicaid programs do possess some monopsonistic purchasing power, and large private insurers may enjoy some degree of monopsony power in some localities, but the highly fragmented buy side of the U.S. health system is relatively weak by international standards. This is one factor, among others, that might explain the relatively high prices paid for health care and for health professionals in the United States.

In comparison, the government-controlled health systems of Canada, Europe, and Japan allocate considerably more market power to the buy side. In each Canadian province, the health insurance plans operated by the provincial governments constitute monopsonies. They purchase (pay for) all of the health services that are covered by the provincial health plan and used by the province's residents. Even pure monopsonists are ultimately constrained by market forces on the supply side—that is, if the buyers offer too little, health care providers will not supply their goods or services. However, monopsonistic buyers may enjoy enough market power to drive down the prices paid for health care and health care inputs fairly close to those reservation prices characterized by the supply curve.

For the United States, most measures of aggregate utilization, such as physician visits per capita and hospital days per capita, typically lag below the medians of other Western countries. Since spending is a product of both the goods and services used and their prices, this implies that U.S. consumers pay much higher prices than consumers elsewhere. However, U.S. policymakers must reflect on what Americans are getting for their greater health spending. Anderson and colleagues conclude that the answer lies in the higher prices paid by U.S. health consumers.

In a follow-up analysis, Anderson and colleagues (2005) revisit the high level of U.S. health expenditures, examining two commonly proposed explanations. The first is that other countries have constrained the supply of health care resources, particularly for elective services, which has led to waiting lists and lower spending. If consumers in other countries must wait for procedures that U.S. consumers can get immediately, then the international consumers are bearing waiting time costs that do not enter national accounts. The researchers argue, however, that the procedures for which waiting lists exist in some countries represent a small part of total health spending. Using U.S. survey data, they calculated the amount of U.S. health spending accounted for by the 15 procedures that account for most of the waiting lists in Australia, Canada, and the United Kingdom. Total spending for these procedures in 2001 was $21.9 billion, or only 3 percent of U.S. health spending in that year.

The second explanation for higher costs is that the threat of malpractice litigation and the resulting defensive medicine in the United States add to malpractice premiums and, more importantly, the practice of defensive medicine, hence increasing costs. Mello and colleagues (2009) update the analysis to address the claim that the U.S. medical liability system leads to unneeded care and extra expenses, examining indemnity payments, administrative costs, and the identifiable hospital and physician costs due to defensive medicine. They estimate these costs to be $55.6 billion, in 2008 dollars, or about 2.4 percent of total health care spending.

The most obvious inference is that eliminating all defensive medicine would have only a minor impact on overall health care spending. Moreover, not all of these costs represent waste—some of them almost certainly provide positive benefits to the patients, or appropriately deter potential malpractice.

Conclusions

In this chapter, we have examined a variety of health care systems found around the world. Variations exist in terms of financing, provider payment mechanisms, and the role of government, including the degree of centralization. The United States stands out with the highest expenditures on health care as well as the highest percentage of the GDP devoted to health care.

Systems that ration their care by government provision or government insurance incur lower per capita costs. In the largely private U.S. system, however, waiting times tend to be shorter than in rationed systems, a conclusion that follows from theory as well as from observation. Americans have been more dissatisfied with their health system than Canadians or Europeans have been with theirs. The study of comparative systems suggests several features of other systems that may be worth adopting. It also suggests that cultural differences among countries could dictate that systems tailored to the local culture continue to differ even in the long run.

Countries have sought to control costs in a variety of ways. Strategies include global budgets, increased cost sharing, and various market incentives. Single-payer plans, as in Canada, offer theoretical economies of administration, but it may be difficult to identify whether the observed cost advantages in Canada would survive translation into a reformed U.S. system.

The United States has fundamentally left cost containment to managed care. Although managed care achieves cost savings and may have contributed to the decline in the U.S. health cost growth rate, its potential will be limited to the extent that employers fail to offer true financial advantages to consumers who choose the low-cost health plans. American-style HMOs, for example, probably would not transfer unchanged to other countries because of cultural and system structure differences.

Most agree that U.S. health system reform must address five critical elements:

- A health "safety net" for all residents, irrespective of age, health status, or employment status.
- Mechanisms that promote cost containment.
- Quality, high-value care.
- Choice for patients and providers.
- Ease in administration.

The next chapter examines these reform elements in more detail, and the degree to which the Affordable Care Act has addressed them.

Summary

1 A useful typology of health benefit systems examines three core dimensions (regulation, financing, and service provision) and three categories of actors (state-based, societal, and private). Combinations of these two sets provide five major systems that summarize most of the more advanced health economies.
2 Among all countries, the United States is by far the biggest spender in absolute per capita terms. It is also the biggest spender as a share of GDP.
3 The United Kingdom's NHS provides relatively easy access to primary and emergency care. It rations elective services either through long waiting lists or by limiting the

availability of new technologies. The NHS devotes considerable resources to high return services, such as prenatal and infant care.

4 The United Kingdom has reformed its health system to include elements of competition. The United States has fewer practicing physicians per capita than the United Kingdom and about the same level of inpatient beds per capita. Health care spending per capita in the United Kingdom, however, is only 37 percent of the U.S. level.

5 The Chinese system has moved from a more "command-based" system to a more market-based system. By 2013, the three insurance schemes covered more than 95 percent of the total population, although benefits vary by insurance scheme due to differing funding levels.

6 The Chinese system has developed a substantive disparity between urban and rural care.

7 The Chinese system has expanded and is now characterized by about one-third of the expenditures covered by the government budget, another third by social insurance, and another third by out-of-pocket payments. Eggleston views the coverage of the current system as "shallow" and "wide."

8 Compared to the U.S. system, the Canadian system has lower costs, more services, universal access to health care without financial barriers, and superior health status. Canadians have longer life expectancies and lower infant mortality rates than do U.S. residents.

9 Canada's single-payer system appears to have substantially lower administrative cost burden than the United States.

10 National health systems appear to reduce health spending. However, careful analysis across alternative systems must impute the additional time costs, as well as differential quality of care in NHS systems, before deciding conclusively on the full costs of alternative systems.

11 Comparisons of health care systems feature competitive systems, as well as varying degrees of monopoly power on the "sell" side of the market and varying degrees of monopsony power on the "buy" side.

12 Monopolistic systems like the U.S. can raise prices above those they would obtain in perfectly competitive markets, thus earning "rents," the excess of prices received by sellers above the minimum prices the sellers would have to be paid to sell into the market.

13 Analysts believe that a monopolistic model characterizes the U.S. system more than systems (Canada, Europe, or Japan) that allocate more market power to the buy side.

Discussion Questions

1 Discuss the factors that may lead one nation to spend more per person on health care than another nation. What are the implications of finding health care to be income elastic in cross-national studies? When health care is income elastic, will richer countries tend to have a higher or lower proportion of GDP spent on health care?

2 In countries in which there is nonprice rationing for care, waiting time costs may be substantial. How could you measure the economic costs of the waiting time?

3 Create a table comparing the British, Chinese, and Canadian health care systems with respect to financing, availability, and costs of care. How do they compare with the system in the United States?

4 Suppose that the price of health care services rises and the quantity demanded falls. Under what conditions might the health care share of GDP fall? Rise?

5 It is important to compare items under the rubric of "all else equal." What crucial factors must be adjusted when comparing health expenditures across countries?

6 Distinguish between a National Health Insurance system and a National Health Service. Provide examples of each. What kind of a system does the United States have?
7 As noted in Table 21.8, patient access and costs vary among a number of measures. Are these measures useful indicators of the performance of health care systems? Could you think of any others?
8 Many economists feel that markets are efficient unless characteristics are present that lead to market failure. What sorts of market failure in the health economy can be used to justify adoption of universal NHI? What kinds of government failure can be used to argue against this proposal?
9 Do countries with more comprehensive national programs for the provision of health care tend to have lower average costs than the United States? Do they have lower rates of growth in costs? Discuss.
10 Speculate about the level of technology available across countries. Do you think that better health care is available in the United States than in Canada? Do international health indices suggest this? What are the complicating issues?
11 What ideas discussed in this chapter would be suitable to recommend to a country just now revising its health system? To pursue equity, that is, wide coverage? To pursue cost containment, that is, lower costs or smaller growth rates?
12 The Organization for Economic Cooperation and Development (OECD) provides some of the best data available for comparative international work. Its website is www.oecd.org. Use the OECD data to examine the health care system of Mexico along the following dimensions:
   - financing
   - expenditure
   - technology
   - coverage.

Exercises

1 Consider the allocation of services in the United Kingdom's NHS, as noted in Figure 21.2. If the government raises the administered price up from P^*, trace what would happen to expenditures in the NHS and in the private sectors.
2 Define income elasticity of health care demand.
   (a) If income increases by 1 percent and the income elasticity of health care demand is +0.75, does the share of income going to health care increase or decrease? Why?
   (b) If income increases by 1 percent and the income elasticity of health care demand is +1.75, does the share of income going to health care increase or decrease? Why?
3 Define price elasticity of health care demand.
   (a) Suppose the price elasticity of health services is –0.4. What will happen to the share of health care expenditures, given a 10 percent decrease in health care prices?
   (b) Suppose the price elasticity of health services is –1.5. What will happen to the share of health care expenditures, given a 10 percent decrease in health care prices?
4 For more advanced students, Table 21.2 provides data for at least rudimentary estimates of income elasticity of health care expenditures. Estimate a regression equation of the following form:

   Log (Expenditures per capita) = $a + b$ log (GDP per capita)

   What is the implied income elasticity of expenditures across countries?

5 From the data in Table 21.2, estimate a regression equation of the following form:

Log (Life expectancy at birth) = $c + d$ log (Expenditures per capita)

What does your resulting equation say about the "effectiveness" of expenditures per capita (without adjusting for any other factors)?

6 Figure 21.3 shows various types of national health insurance systems.
   (a) Compare the total expenditures in panels A and B. Which set of expenditures is larger? What determines which will be larger? Why?
   (b) Compare the total expenditures in panels B and C. Which set of expenditures is larger? In which are resource costs larger? Why?
   (c) If the demand curves truly reflect consumer preferences, which of the three panels is economically efficient? Show the economic losses and the transfers for those panels that are not economically efficient.

Notes

1 The Social Security Administration maintains an archive called *Social Security Statistics throughout the World*, describing the health, unemployment, and poverty programs in more than 170 countries, at www.ssa.gov/policy/docs/progdesc/ssptw/, accessed February 16, 2016.

2 Updates and summaries are from Allin and Rudoler (2015).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 22

Health System Reform

In this chapter

- Goals of Reform
- Ensuring Access to Care
- Quality of Care
- The Affordable Care Act (ACA) of 2010
- Competitive Strategies
- ACA Outcomes after Six Years
- Meeting Reform Goals
- Conclusions

Chapter 21 examined national health insurance (NHI) programs in several major industrialized countries. The United States established a more comprehensive health care system in 2010 with the Affordable Care Act (ACA), but many issues remain. According to the nonpartisan Kaiser Family Foundation, in 2014 about 32 million Americans did not have health insurance, a decrease of 9 million people since the previous year, and about 18 million people since 2011. While many still favor universal health care coverage as a solution to the problems of access and costs, others argue that we can meet health care objectives more effectively through reforms that reduce the role of government and instead take advantage of market forces. The basic issues in health system reform and alternative reform proposals are the focus of this chapter. We finish the chapter with an extensive presentation and evaluation of the ACA.

Goals of Reform

Most would agree that a national health system reform must address these five elements:

- A health "safety net" for all residents, irrespective of age, health, or employment status.
- Mechanisms that promote cost containment.
- Mechanisms that promote quality and high-value care.
- Choice for patients and providers.
- Ease in administration.

Consider the five elements in order for the United States:

Safety net—Large portions of the U.S. population receive inadequate health care by almost any criteria. While Medicare provides almost universal health care for those over age 65 and Medicaid/CHIP are making great inroads into the population under age 18, millions of Americans still lack access to levels of health care that even the most conservative analysts would view as adequate.

Cost containment—The United States spends well over one in six dollars of its GDP on health care, and expanded coverage will almost certainly increase that ratio. While some analysts have argued that this amount may reflect consumer preferences for high-quality health care, there are clearly potential reforms relating to administrative costs, and ineffective treatment, that could reduce overall health care costs. Further, few Americans would desire cost containment at the expense of the quality of the health care.

Quality, high-value care—Improved access to care and cost containment have been long-standing goals of health system reform. There is a growing consensus in the United States and elsewhere that reform efforts should also promote high-value, cost-effective care.

Choice for patients and providers—Contrasting the failure of President Bill Clinton's 1993–1994 reform initiative with the passage of President Barak Obama's ACA in 2010 suggests that successful U.S. reform must provide choices of providers and treatments.

Ease in administration—Consider the weekly trip to the supermarket. The decisions on where to shop and what to buy, while constrained by budgets and the prices of the goods, are administratively simple. People go where they shop, buy what they need, and need not deal with bureaucrats or forms. Even those who receive government benefits to buy food get easy-to-use "bridge cards" to pay at grocery store cash registers. Contrast that to the U.S. health insurance systems, with different application

forms, insurance forms, cards, and a myriad of questions about who pays for what, and whether what one has paid will be reimbursed. While purchasing health care is obviously more complicated than purchasing food, any national health care policy that simplifies the process would be desirable.

Many reform proposals face the dilemma whether to fund coverage by *individual* mandate, *employer–employee* mandate, or *general revenues*. An individual mandate is a law that requires individuals to buy health insurance for themselves, with subsidies, usually funded out of general revenues, for those who cannot afford it. Employer–employee mandates would require taxes on wages for the employee's share. The employer's share may also fall on the employee in the form of lower wages. Subsidies out of general revenues would provide for the unemployed.

Advocates of a Canadian-style system seek a single-payer system with government revenues providing most of the NHI. Yet another alternative is the medical savings accounts method of payment that would allow people to set up a tax-free savings account out of which they can pay the out-of-pocket costs of their health care. Usually, these systems entail health insurance with high deductibles and coinsurance.

Some plans define a minimum acceptable insurance level, with people allowed to purchase more extensive coverage if desired and if they can afford it. Others define one plan to fit all. In early U.S. debates, many plans featured universal coverage. As support for universal coverage waned in Congress, proposals for NHI sought less ambitious goals.

Basic Issues in Reform

Any reform program must face difficult questions, one of which is *how much* service coverage. Clearly, covering more services or mandating a larger variety will increase costs.

Figure 22.1 shows an economy that allocates its resources to nonhealth and health goods (and services) at Point *A*. It would be best, of course, if *A* were on the production possibility frontier of efficient production for nonhealth and health goods, the solid line PP', but there are many reasons that it is probably not. Ineffective treatments, needless tests, and excessive paperwork may all provide less health (and other goods) than possible, so that we see an interior frontier indicated by the dashed line PP''. For simplicity on the dashed curve we have drawn the production of all other goods as efficient at point *P* (on curve PP''), although there is no reason to believe that other goods are produced more (or less) efficiently than health goods.

Assume that the society determines to provide a safety net for all residents, increasing the amount of health goods provided from H_0 to H_1. The *economic cost* of providing $\Delta H = H_1 - H_0$ of health goods is the amount of *G* given up, or $\Delta G = G_0 - G_1$ at point *B*. If we could control costs, or provide health goods more efficiently, society might plausibly reach a point like B' or B'', on the efficient frontier. One underlying goal of reform would be to move to a more efficient production of health from health goods. A related issue is whether there will be cost sharing for covered services and, if so, what type of cost-sharing arrangement will occur.

The question of *who* to cover can be equally difficult to address. On any given day in the United States, there are millions of foreign students, visitors, and temporary workers as well as millions of illegal immigrants. The difficulty of determining the covered population in some cases is evidenced by the acrimonious debate over care provided (and paid by governments) to undocumented workers and their families.

A third major issue is how to fund health reform. Will it rely on general tax revenues or will funding come from mandates on businesses and/or individuals? In either case, where will the burden of funding ultimately rest?

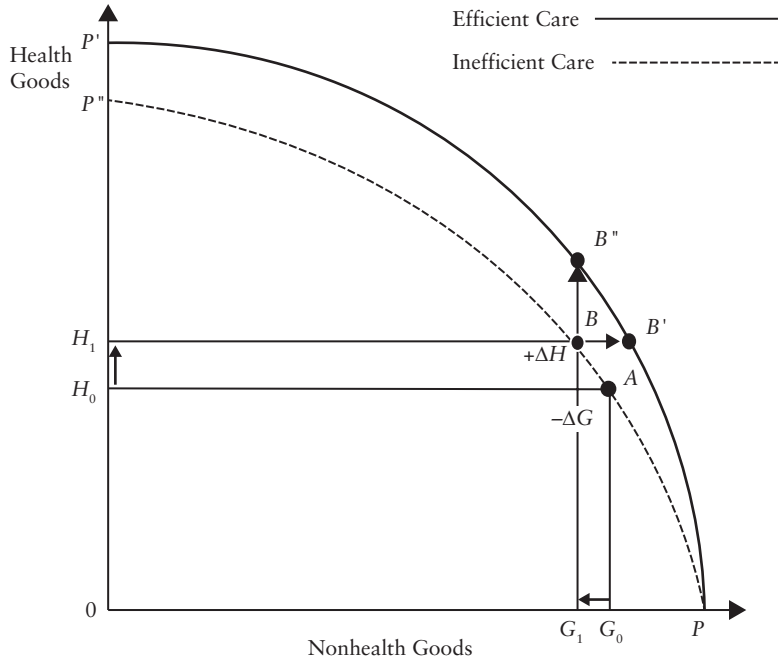


Figure 22.1 The Resource Costs of Health Reform

The most challenging issue is to determine whether health reform will build largely on the existing framework of government programs and private employment-based insurance with most of the reform effort aimed at cost containment and reducing the pool of uninsured. Other proposals, especially market-oriented proposals, attempt to attain these goals by weakening the link between private insurance and employment.

The Costs of Universal Coverage

NHI programs that guarantee universal coverage certainly cost a great deal of money, but it is important to distinguish which costs are incremental. In other words, what are the *additional* costs to society from the imposition of NHI?

From society's point of view, the incremental cost of NHI in the United States is the extra expenditure on health care incurred if we switched to national health insurance. Inasmuch as most people already have insurance for almost all hospital care and most physician care, the extra cost of NHI would be smaller than many expect.

One reason is that the uninsured already consume health care. Zero insurance does not necessarily mean zero care. The major reason for switching to an NHI plan is to extend coverage to the uninsured (50 million in 2010; 41 million in 2013; 32 million in 2014).

Coughlin and colleagues (2014) estimate the expenditures of those who are uninsured or partially insured as shown in Figure 22.2. Relying on pre-ACA data from 2008 to 2010, they estimate that in 2013, the average full-year uninsured person (40.8 million) had half the medical spending of the average insured person ($2,443 versus $4,876). The average part-year insured person (31.4 million) spent $3,439. Assuming that a universal plan would

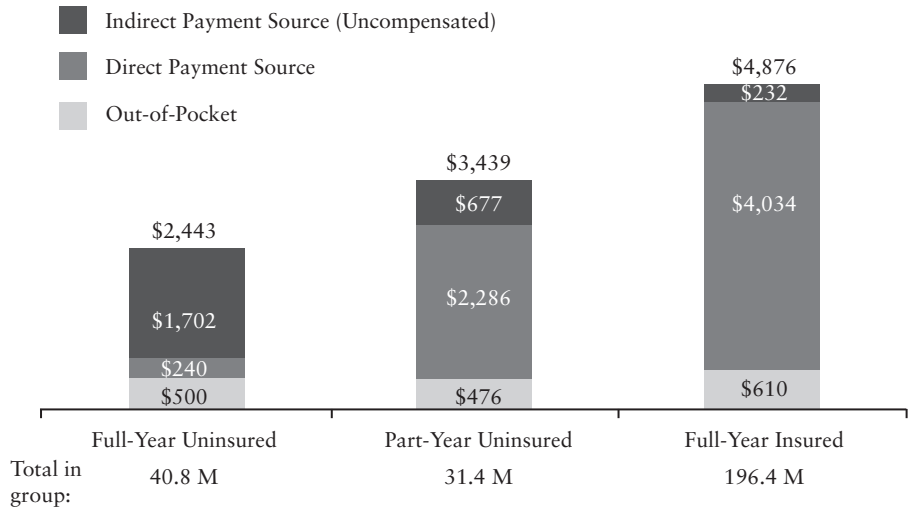


Figure 22.2 Per Capita Medical Spending among the Nonelderly

Note: "Direct payment source" among the full-year uninsured includes "other public" payments, which are Medicaid payments. These payments may be retroactive or emergency provided by Medicaid.

Source: Medical Expenditure Panel Survey—Household Component, 2013–2014

provide the same compensation to all, a crude calculation raising the level of expenditures from the lower levels to the higher level for the 72.2 million uninsured would give an additional expenditure of $144.4 billion per year, or roughly 5 percent of national health expenditures (NHE).

The incremental cost will be higher to the extent that a national plan provides greater typical coverage than people already choose to buy or have provided to them by other sources. Also, any tax-supported system of financing care potentially entails a deadweight loss to society, as taxpayers respond to higher taxes by working or investing less, leading to some efficiency loss. This is true even if the program is of the employer-mandated type, because a law forcing employers to incur expense is really a tax.

The incremental costs constitute real costs to society, because as shown in Figure 22.1, we must divert resources from elsewhere to pay these costs. In contrast, differences in financing methods (determining who pays) mean less in economic terms. Policymakers may find it more palatable to choose a plan that does not greatly expand the government budget, and they may choose employment-mandated plans for this reason. Nonetheless, society incurs the cost irrespective of whether it finances it through the government or through mandates to individuals or employers.

Ensuring Access to Care

In this section, we group reforms by their two main motivations: the desire to see that people get needed health care, and the desire to control the rising cost of health care.

Employer versus Individual Mandates

The country that wishes to provide universal coverage for health care must choose one scheme or another to extract resources from its households. Schemes for employers or government to pay the bills are only mechanisms by which households ultimately pay. The U.S. debate features and contrasts two mechanisms: employer versus individual mandates.

Under *employer mandates*, employers must procure health insurance for their employees and their dependents. Employer mandates form the backbone of the health systems in Europe, Latin America, and Asia (Krueger and Reinhardt, 1994). Although the employers write the checks, competitive firms undoubtedly will pass on as much of this cost as they can to customers as higher prices or to employees as lower wages. The *individual mandate*, in contrast, obligates all residents to purchase health insurance for themselves and their families, either from private insurance (individually purchased) or through a group, such as a work group, professional organization, or religious group. The government subsidizes the poor in their purchases by taxing those who have more money.

In Chapter 11, we showed that a lower market money wage rate leads an employer to hire more workers. Assuming at the outset that there are no health benefits and that the market wage is $20 per hour, employers will hire workers as long as the marginal revenue from the goods those workers produce exceeds the $20 per hour wage. Suppose, to begin, the employer hires 1,000 workers. Suppose also that an NHI requires employers to provide a health benefit for all workers that costs $1 per hour of work. If the mandated benefit is worth at least $1 per hour to the workers, and costs exactly $1 per hour for employers to provide, those employers who were previously willing to pay $20 will now pay $20 less the $1 cost to provide the mandated benefit. Other points on the demand schedule will also decrease by the $1 cost of the benefit.

Workers previously willing to accept a wage of $20 will now be willing to supply their labor for $1 less since they value the mandated benefit at $1. The net wage (money wage + the value of the benefit) remains unchanged at $20, but the equilibrium money wage falls to $19, or by exactly the amount of the benefit. Workers accept lower money wages, and the same 1,000 workers are employed at the same net wage, $19 in money wages plus the $1 benefit. The workers are no worse off at a wage of $19 with the mandated benefit than at $20 without the mandated benefit because the benefit is worth the $1 that it cost in reduced wages.

Business leaders often complain that employer mandates either will reduce profits or force firms out of business. Such responses implicitly assume that their firm is the only one affected by the mandate. If all firms faced the same labor costs, it is doubtful that closings would result. In the short run, firms would pay workers less, take less in profits, and/or raise prices to consumers. "Economists are convinced, however, that in the longer run more and more of the cost of the employer mandate would likely be shifted backward to employees . . . through smaller real (inflation-adjusted) increases in wages than would have been warranted by long-run productivity gain" (Krueger and Reinhardt, 1994, p. 44).

If the labor supply is very unresponsive to the wage rate, or *inelastic*, the employer's lower wage expenditures will offset extra health benefit costs regardless of whether the laborers value the benefit highly or not at all. Most economists would agree that the aggregate labor supply, at least in the long run, is nearly vertical (totally inelastic) for men, and also highly inelastic for women. In this scenario, the mandate has little effect on producers, their competitive positions, either domestically or internationally, or their customers. Whether the program helps or harms the society's well-being under conventional economic analysis depends largely on whether workers value their health insurance as much as or more than they did their foregone wages.

The *individual mandate* provides the same result with a clearer pathway, because the costs fall on the beneficiary who pays them directly. Pauly (1994b, 1997) describes an individual mandate, enforced by employers and subsidized for the poor, requiring all individuals to purchase a minimum health plan or better. He argues that this approach is desirable so that people can relate their taxes to what they are paying to obtain benefits. In this scheme, individuals must purchase health insurance. They may in fact acquire it through their workplaces, or they may buy it explicitly in a market setting.

During the U.S. debate over President Bill Clinton's 1993–1994 proposals, disputes frequently arose over the fraction that the employer pays as opposed to the fraction paid by the individual, on the presumption that the chosen fraction reflects the burden. Economists, however, tend to agree that the fraction chosen does not matter. The discussion presented above (regarding the $20 per hour wage) says nothing about fractions. The economic logic suggests that those who are least able to avoid a tax will bear its burden, irrespective of who writes the check. Some argue that it is politically necessary to overlook the economics, but others insist that an open public discussion of the genuine issues would improve the quality of national debate.

Separation of Health Insurance from Employment

Those seeking to redesign a health system can make a good argument for revising or replacing the prevailing system of employer-provided insurance with either a single-payer system or an individual mandate. The advantages of employer provision stem from long-term practices that cause economic distortions. During World War II, the U.S. government froze prices and wages. Competing for workers, firms expanded their fringe benefits, which were not subject to the freeze. After World War II, employer contributions to health insurance were, and continue to be, tax-exempt, providing workers with a substantial discount and inviting inefficiencies of over-insurance and moral hazard. Meanwhile, even under the ACA, many unemployed, as well as many low-wage employed, have gone without health insurance.

Health insurance problems also occur when workers change jobs. When leaving their previous employer's health coverage behind them, workers have little choice but to buy an individual policy, a "continuation of benefits" or COBRA, from the previous employer, or do without insurance entirely. Individual policies are often more expensive, sometimes pose administrative problems, and sometimes comprise a lower financial priority for people out of work. Before the enactment of the ACA, workers often found pre-existing conditions such as heart disease to be uninsurable.

Single-Payer versus Multiple Insurers

A move in the United States toward universal coverage also entails the option of a single insurer, presumably the federal government. In the United States, multiple private companies insure a majority of the population. The most prominent single-payer proposal discussed in the U.S. debates has been the Canadian NHI system (known in Canada as Medicare).

Economic theory suggests that consumers value variety. Within any given city, numerous restaurants serve different foods, prepared in different ways. American auto manufacturer Henry Ford said (at least apocryphally) that buyers could have any color of his pioneering Model T, so long as it was black—his company lost its market prominence to General Motors which provided a wider variety of cars (and colors). In principle, a variety of insurers may provide different coverages, pool different groups, and create products that more closely match the variety of consumers.

However, the U.S. multiple-insurer system has led to multiple forms and policy rules that face hospitals, clinics, and nursing homes. Patients, as insurance clients, must often provide the same information numerous times, with commensurate possibilities of error. With hundreds of different health insurers, the difficulty of coordinating different policies falls on hospitals, physicians, and ultimately the policy-holders themselves. These coordination problems are external to the insurance companies, who do not see the need to reduce them. Nevertheless, coordination of policies and coverages constitute real economic costs both to patients and to providers. The government as single payer could reduce those costs with fewer and standardized forms, or electronic chips on cards to provide machine-readable, and more accurate, information.

In principle, consolidation of insurers could reduce administrative costs if there are economies of scale in administration, or if gains could obtain from pooling those insured. Many economists have tried to estimate the excess administrative costs. Cutler and Ly (2011) partition the $1,589 difference in per capita health care spending between the United States and Canada in 2002. Higher administrative costs accounted for $616, or 39 percent, of the difference. The authors argue that this figure probably underestimates the amount and share, because nurses also spend substantial time on administrative tasks, but accounts typically consider nursing time as clinical care rather than administration.

Could a government single-payer system solve all of the administrative cost problems? The same administrative technology is available to the private sector, and if further economies were possible, and there is appropriate nonmonopoly competition, private firms could profitably merge to provide cost-reducing service. One must also note that private insurer profits are not a waste to the economy, but rather payments for capital that government also must incur.

Moreover, a switch to a single-payer system would greatly diminish the very large private health insurance industry. To put the issue in perspective, in 2014 private premiums were $991 billion or just about 33 percent of total national health expenditures. To be blunt, one in three dollars of health care expenditures goes through private insurance! Private insurers would almost certainly oppose a single-payer plan, and they have actively promoted their own interests in the formulation of the ACA.

Do health care system problems warrant a change to a single payer? Other reforms may address specific problems. For example, we insure the uninsured through mandated coverage including subsidies for the poor, and provide coverage for people with pre-existing conditions through the individual mandates. Before the ACA, workers who were previously insured were allowed to purchase their previous coverage for a limited time from their previous employers.

A potential benefit of the single-payer system lies with the possibility of common coverage. We may worry now that some insured people have inadequate policies in terms of the depth and breadth of coverage. The single payer could offer one policy or a small number of variations, with each variation determined to be adequate by policymakers and interest groups representing the public. In fact, all Americans ages 65 and over who participate in Medicare (just about all of them) face similar basic choices. In contrast, the availability of many policies from many companies, while offering variety and tailoring policies to the individual preferences for cost-sharing features and coverage, makes the policies difficult to compare.

Quality of Care

Improving the quality of care has become a vital component of health system reform. In the presidential campaign of 2008 prior to passage of the ACA, both major party candidates stressed the need for higher health care quality. Senator John McCain vowed “to make

sure they [patients] get the high-quality coverage they need." Then-Senator Barack Obama devoted an entire section of his proposal to the goal of "ensuring providers deliver quality care." The ACA legislation that was ultimately signed into law included numerous provisions designed to improve quality and patient safety, and the major ones will be described later in this chapter.

Previous chapters have described clinical studies that detail major gaps and unevenness in quality, and mechanisms such as health care report cards to encourage quality improvements by embarrassing providers who perform poorly and by influencing patient choices. We have also described the use of pay-for-performance (P4P) incentives to raise quality.

Health economists continue to grapple with quality issues, expressing concern about:

1 Moral hazard and the overutilization associated with insurance (a theme we have stressed throughout the text).
2 The insufficient consideration of cost-effectiveness analyses to distinguish economically efficient from inefficient procedures, technology, and levels of care (see especially Chapter 4 and Box 22.1).
3 The limited use of financial incentives to promote quality.

Giving greater priority to financial incentives is a theme that we have also stressed throughout this text. Restructuring incentives appears straightforward, in principle, but the transition from concept to practice is often very challenging. While the success or failure in healing a broken bone is relatively easy to monitor, quality aspects for many conditions are much more difficult to define and monitor, particularly for chronic conditions. Paying for performance requires sophisticated definition of performance measures, and determination of the appropriate incentive amounts needed to influence provider behavior.

BOX 22.1

Preventive Care and Cost-Effectiveness Analyses

A greater emphasis on preventive health has become a mantra for many political and thought leaders. As we shall note later, it has also become an important element of the ACA. Diabetes screening for type 2 (adult-onset diabetes) is one prominent example. The health-related consequences that arise from this disease are staggering. Are widespread screening efforts for this disease cost-effective? Cohen and colleagues (2008) urge caution against sweeping generalization regarding preventive care.

Consider just the following examples of preventive measures they provide (taken from the Tufts–New England Medical Center Cost-Effectiveness Registry). The incremental cost per quality-adjusted life-year (QALY) is in 2006 dollars.

High-intensity smoking relapse program (compared to low-intensity program)	$190
Intensive tobacco use prevention program for 7th and 8th graders	$23,000
Screening all 65-year-olds for diabetes (compared to diabetes screening of all 65-year-olds who have hypertension)	$590,000

By almost any standard (typically $50,000 to $100,000 per QALY), the first two prevention programs meet the threshold for adoption. The third, unlimited screening of all 65-year-olds for diabetes, should unequivocally be rejected.

Two important lessons emerge. First, we cannot make rational decisions without reliable cost-effectiveness values. Second, policymakers and third-party payers must discriminate carefully within preventive (and undoubtedly other) categories of health care interventions. Political messages that sound good can reflect bad economics.

P4P is now common in both private and public insurance plans, with mixed evaluations. Pearson et al. (2008) evaluated P4P programs introduced by Massachusetts' leading commercial insurers. The research covered a wide variety of P4P contacts with a large number of physician groups over the period 2001–2003. It showed that the quality improvement, represented by 13 HEDIS measures, for highly incentivized groups was not larger than the improvement found in comparison groups.[1]

In contrast, the Tax Relief and Healthcare Act of 2006 mandated a P4P program for Medicare. The program, known as the Physician Quality Reporting System, is still voluntary, but participation has grown rapidly from 55,000 professionals in 2007 to 585,000 professionals within 45,000 practices in 2014. The CMS paid an average bonus of nearly $5,000 per participating professional practice in 2014.

The Affordable Care Act (ACA) of 2010

The U.S. Congress passed the ACA in March 2010. The primary goal was to reduce the number of uninsured people in the United States, then close to 50 million, while maintaining a viable private insurance system. Here we will:

1 Discuss the logic behind the so-called "three-legged stool" analogy that characterizes the ACA.
2 Touch on the major features of the (very complex) Act.
3 Provide economic analysis of key features of the Act.

The following section will evaluate the outcomes six years (as of 2016) after passage.

The "Three-Legged Stool"

Jonathan Gruber, one of the ACA's architects, and a key participant in the earlier Massachusetts Health Care reforms of 2006 (often referred to as Romneycare, after then-Governor Mitt Romney), uses the analogy of the three-legged stool (Gruber, 2010) to characterize the three features of the ACA. Geometrically, a stool needs no fewer than three legs to establish a plane, stand stably, and bear the weight of considerable use.

The first "leg" is the requirement that insurance companies offer insurance to any applicant with premiums based on age (and tobacco use) and *not* on underlying health status. Insurance companies may *not* exclude applicants due to pre-existing illnesses. While in principle any event is insurable, actuarially fair health insurance policies may be prohibitively expensive to individuals or groups of patients. Without further requirements, healthy patients may choose to exit any group that is charging these actuarially fair premiums. The second leg addresses the market impacts of the universal coverage mandate, and the third leg addresses the individual impacts.

As we noted in Chapter 10, if insurance companies charge the same price to people whether they are sick or healthy, many healthy people may see this as a "bad deal" and not buy insurance. Their exit results in higher prices that chase even more people out of the market, and can lead to the destruction of the markets. To combat this possibility, Massachusetts in 2006 added a *second "leg"* to the stool by requiring that all residents carry insurance. In this way the state could ensure a broad distribution of health risks in the market and fair "community-rated" pricing to all. Gruber couches this requirement in insurance terms, but a large public finance literature supports the affluent paying (increased) taxes, to fund transfers to the less affluent (or their widows or children) to achieve economic equity. This, in fact, was the rationale that the Supreme Court used in upholding the ACA in 2012.

The first two legs establish and maintain insurance markets, but they do not guarantee that buyers can afford the premiums that would allow insurers to stay in the market. The ACA therefore added a *third "leg"* in the form of subsidies that make health insurance affordable for those whose incomes would not otherwise allow them to buy it. Although the rules are somewhat complicated, the ACA introduced two types of subsidies—*tax credits* and *cost sharing*. Tax credits reduce the costs to consumers of paying premiums, while cost sharing reduces their out-of-pocket costs by lowering copayments, deductibles, and the maximum out-of-pocket costs that can be assessed over the policy period. For 2016, for example, the premium tax credits that reduce the prices of insurance premiums apply to people between 100 and 400 percent of the federal poverty level (FPL). The *cost-sharing* reductions apply to those between 100 and 250 percent of the FPL.

The ACA—Basics

Legislation has words, rather than "legs" of a stool. The ACA requires that most U.S. citizens and legal residents have health insurance. It creates health insurance marketplaces, commonly (and henceforth) referred to as "health exchanges." Individuals or families can purchase coverage through these exchanges, with premium and cost-sharing credits available to those with incomes between 138 and 400 percent of the federal poverty level (the poverty level was $20,160 for a family of three in 2016). The ACA also created separate exchanges through which small businesses can purchase coverage.

The ACA requires that employers pay penalties for employees who receive tax credits for health insurance through an exchange, with exceptions for small employers. It imposes new regulations on health plans in the exchanges and in the individual and small group markets. Finally, starting in 2014, states participating in Medicaid expansion could increase eligibility levels within their state to 138 percent of the Federal Poverty Level (about $16,400 for an individual and $33,500 for a family of four in 2016).

Table 22.1 Essential Benefits under the Affordable Care Act

1 Ambulatory patient services.
2 Emergency services.
3 Hospitalization.
4 Maternity and newborn care.
5 Mental health and substance use disorder services, including behavioral health treatment.

Table 22.1 *continued*

6	Prescription drugs.
7	Rehabilitative and habilitative services and devices (services and devices that help people keep, learn, or improve skills and functioning for daily living).
8	Laboratory services.
9	Preventive and wellness services and chronic disease management.
10	Pediatric services, including oral and vision care.

At the benefits level, the ACA assures an "Essential Benefits" package summarized in Table 22.1. It includes some very familiar services such as emergency services and maternity and newborn care. It also includes some less familiar services such as rehabilitative and habilitative services and devices (related to daily living). Examples include therapy for a child who isn't walking or talking at the expected age. The ACA imposes annual out-of-pocket (OOP) maximums on the amount that enrollees in most health plans—including self-insured and large-group health plans—must pay for covered essential health benefits through cost sharing.

Some analysts viewed the exchanges as a type of Orbitz® (a travel website) for buying health insurance. Recognizing that more inclusive plans would cost consumers more, the ACA created four general benefit categories, referred to by *metals* platinum, gold, silver, and bronze.

Platinum plans provide the essential health benefits, 90 percent of the benefit costs, with an out-of-pocket limit equal to the Health Savings Account (HSA) limit (for 2017, equal to $7,150 for an individual, and $14,300 for a family). Gold plans provide the essential health benefits, covering 80 percent of the HSA out-of-pocket (OOP) limits. Silver plans provide 70 percent and bronze plans 60 percent, again with the same OOP and HSA criteria. Catastrophic plans pay less than 60 percent of the total cost of care, with consumers paying the balance. These plans are only available to people less than 30 years old at the beginning of the plan year, or those with a hardship or affordability exemption. The OOP limits are decreased (lower levels for lower incomes) for those up to 400 percent ($47,080 for a household of one in 2016) of the federal poverty limit.

Economic Analysis of the ACA

Two important parts of the ACA merit a graphical analysis. The Supreme Court's 2012 affirmation of the ACA allowed states to refuse to opt into the Medicaid expansion. We saw in Chapter 20 that typical Medicaid matches outside of the ACA are on the order of 50 percent, reducing the price of health care relative to all other goods by about one-third. As seen in Figure 22.3 the initial ACA matches were on the order of 8 to 9 times, reducing the price of Medicaid care by about 90 percent, moving from point X_{Pre} to X_{Post}. These matches could fund substantial health care increases in the states. While some states might potentially have to raise taxes or re-allocate funds to receive the matches, the foregone gain to large numbers of residents is so large as to make states' refusals implausible on economic grounds. The states must be using other reasons for refusing to opt into the program.

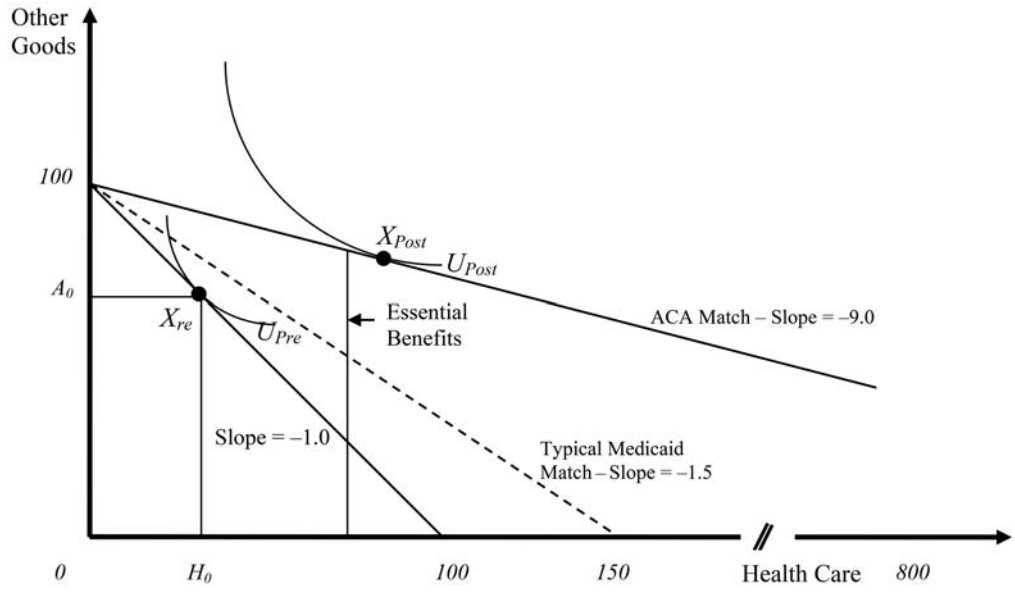


Figure 22.3 The Impact of the Medicaid Expansion

A second analysis involves the availability of heavily subsidized insurance on the choices that consumers make. In Figure 22.4 we see that initially Bob's family purchased no insurance (point *B*), while Dave's family purchased a small policy (point *D*). The availability of low cost, or subsidized, insurance at point *I*, leads both to purchase the insurance. In terms of Chapter 20, Bob's family "takes up" the insurance and they now have more insurance than before (when they had zero). Dave's previous insurance is "crowded out" by the new insurance. We see, however, that in this example Dave's family has chosen a smaller policy than before.

Instances of behavior like Dave's (in this example) made headlines, particularly among ACA opponents, but the economics are clear. Dave's family is taking a smaller policy, but at a much lower cost, freeing up funds for other items that it values more. This effect is well-known in the analysis of housing vouchers, for example, where households may choose a smaller unit because it costs much less. Rather than representing a program failure, it represents the opportunity for the program to tailor consumption more specifically to households' needs.

Despite its focus on the uninsured, the ACA affects almost every segment of the health economy. For example, numerous provisions affect Medicare beneficiaries as well as those who provide services to them. In previous chapters, we have described reductions in payments to Medicare Advantage Plans (Medicare Part C) and the gradual elimination of the "doughnut hole" by 2020. Other provisions call for reductions in payments to hospitals and other providers.

To raise revenues and discourage over-insurance, the Act imposes a "Cadillac" tax on high-cost employer-provided policies (recall that the U.S. tax system gives disproportionate subsidies to high-cost policies through the graduated income tax deduction). In addition to this tax, funding for the program comes mainly from an expanded Medicare tax base that will affect higher-income individuals and families, fees on health insurers, and taxes on

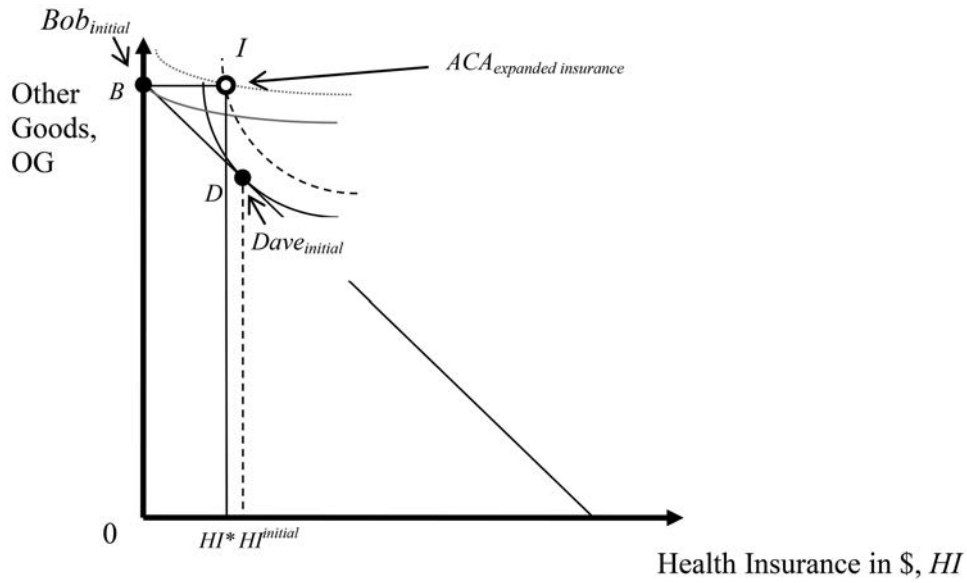


Figure 22.4 The Household Impact of Exchange-Provided Insurance

manufacturers and importers of branded drugs and certain medical devices. These new revenue streams will nevertheless fall short of the increased federal obligations under the ACA.

To close the gap, the Obama administration and supporters of the ACA have emphasized measures to "bend the curve," that is, slow down the overall rate of growth of health care spending through increased preventive health, administrative simplification, and by reducing inappropriate care. To accomplish the latter, the ACA has created an independent, non-profit Patient-Centered Outcomes Research Institute. Although the Institute has only limited authority in making recommendations, it examines and conducts studies comparing the appropriateness of alternative treatments.

Competitive Strategies in the Post-ACA Era

Until passage of the ACA, an ideological battle raged over the superiority of (1) increased government involvement through both expanded regulation and additional government programs to provide or finance health care or (2) an increased emphasis on market mechanisms and market forces with corresponding decreases in the use of regulatory instruments. The Republican Party platform, adopted for the 2016 presidential contest, demanded that the ACA:

> must be removed and replaced with an approach based on genuine competition, patient choice, excellent care, wellness, and timely access to treatment. To that end, a Republican president, on the first day in office, will use legitimate waiver authority under the law to halt its advance and then, with the unanimous support of Congressional Republicans, will sign its repeal.[2]

The Republican Party, however, had not offered a cohesive plan to replace "Obamacare." A careful examination of the ACA and other developments in the post-ACA period will reveal that promoting competition and innovation remains central to U.S. health care policy.

The ACA is designed to promote competition through at least two mechanisms. First, federal subsidies for insurance are given only to those who purchase coverage through the exchanges that we have described. Second, the exchanges make it much easier for consumers to compare costs of standardized policies. Thus, to gain access to consumers, insurers have incentives to participate in the exchanges, and if enough insurers do, to offer competitively priced policies. It is still too early to determine whether this strategy is working as intended.

We note one final market approach known as *Value-Based Insurance Design* (VBID) that relies on financial incentives. VBID, established in 2005, preceded the ACA, but one part of the law created a new section of the Public Health Service Act that features VBID coverage. Under VBID, patient copayments are reduced for high-value care and its follow-up, and raised for lower-value services. These programs are often directed at preventive and chronic care. VBID is currently under wider consideration, and empirical evidence on its effects has become available. In reviewing the literature, Lee and colleagues (2013) concluded that VBID has improved the quality of care but has not reduced overall spending. Elsewhere, Hirth et al. (2016) compared participants in a voluntary VBID program for Connecticut state employees with state employees in six other states. Compared to the other states, Connecticut employees in the first two years of the program had greater use of targeted services and better adherence to medications for chronic conditions. Cost comparisons did not find any significant differences.

Development of Alternative Delivery Systems

The dominant competitive strategy in the United States has been to promote delivery systems that can provide alternatives to traditional fee-for-service with its comprehensive first-dollar insurance coverage. The cornerstone of this strategy has been the promotion of various forms of managed care, including health maintenance organizations (HMOs) and preferred provider organizations (PPOs).

Managed care health plans, described in Chapter 12, provide comprehensive sets of services for fixed monthly premiums. They typically feature minimal cost-sharing for covered services. In exchange for expanded coverage, patients' choices are restricted to specific providers and hospitals. Furthermore, the patients' primary care physicians serve as the gatekeepers for referral to most specialists and nonemergency hospital admissions.

Expansion of the population covered by managed care has been a major policy goal. The primary motive behind the managed care strategy is the view among many policymakers that the traditional fee-for-service form of health care delivery was the primary cause of rising costs and unnecessary care.

Managed care has reduced costs through lower hospitalization rates (inpatient care) and lower payments to providers. However, managed care cost increases have paralleled rates of increase elsewhere, limiting the overall potential for cost containment. Nevertheless, federal policy continues to emphasize the managed care strategy by proposing expanded incentives for those who select prepaid, managed care systems, especially Medicare and Medicaid beneficiaries.

The ACA has also promoted Accountable Care Organizations (ACOs). ACOs, as described in Chapter 12, are integrated entities that involve tight coordination of care coupled with joint financial incentives to providers. The development of ACOs preceded the ACA but the ACA

legislation led to the "Next Generation ACO Model." ACOs under this program assume greater performance risk while potentially receiving a larger share of savings than other types of ACOs.

Consumer-Directed Health Plans and Health Savings Accounts

Consumer-directed health plans (CDHPs) paired with health savings accounts (HSAs) represent an important health care delivery alternative. Proposals for medical savings accounts go back to the early 1980s (Stano, 1981). Under proposals that emerged in the following years, employers or public payers would contribute to an individual's *Medical Savings Account* (MSA). The account would allow the holder to purchase relatively low-cost catastrophic insurance with high deductibles. Holders would then use MSA balances to pay out-of-pocket costs due to the deductibles, while providing true catastrophic insurance for large unexpected charges. The account would then distribute the unused portion in the MSA at the end of a designated period or at retirement.

MSA advocates contrast it to comprehensive, tax-subsidized insurance, which creates substantial moral hazard and ineffective incentives for efficient consumption of care. They argue that potential distributions from an MSA, like spending their own dollars, provide individuals with incentives to use care prudently. In principle, patients will be less likely to consume unnecessary or marginally beneficial care, and the stronger market forces will help restrain prices.

After a restrictive MSA form (meaning that it was completely controlled by the employer) was approved in 2002, a tax-advantaged health saving account (HSA) legislation passed in 2003. The HSA is a less restrictive MSA, owned by the *employee*, and open to anyone enrolled in a high-deductible health plan and not already covered by public or private insurance. In 2016, the minimum deductible had to be $1,300 ($2,600 for families). Individuals with qualified coverage were allowed to contribute up to $3,350 ($6,750 for families) to their HSAs.

The motive for this CDHP strategy is to create highly informed consumers and to give them the incentives and the tools so that they take charge of their health care decisions. Their search for price and quality would counter the power of medical providers and the inefficiencies in the current marketplace. In this sense, the CDHP contrasts sharply with managed care under which the patient is a more passive participant and where the managed care plan administrators take responsibility for prices and quality. As noted in Chapter 12, 24 percent of covered employees participated in high-deductible plans in 2015. In 2007, Medicare also introduced an MSA option but enrollments are typically less than 3 percent of Medicare beneficiaries.

Feldman and colleagues (2007) analyzed a three-year window for plans offered by a large employer and find little significant savings for those enrolled in CDHPs. Dixon et al. (2008) examined a large manufacturing company that had large- and low-deductible CDHP options, with premiums lower in the former, as well as a PPO plan. Enrollees in the high-deductible CDHP were more likely to cut back on utilization, but they were more likely to engage in risky cost-saving behavior (e.g., not going to a physician when they should have, or taking less than the recommended dose of a prescription drug). Another report for the same firm found that the high-deductible CDHP enrollees were substantially more likely to discontinue taking some categories of drugs used to treat chronic conditions (Greene et al., 2008b).

Analysts temper their enthusiasm for CDHPs by recognizing the undesired consequences that may result from a system that depends on voluntary enrollments. Healthier individuals will more likely choose high-deductible health plans. They may purchase catastrophic coverage at relatively low rates and will more likely have funds left over in their HSA accounts. This selection phenomenon might also lead to escalating premiums for the sicker populations who remain in managed care and conventional plans so that the net effect could turn out to be largely a

redistribution of income toward the healthy. Clancy and Gauthier (2004) provide an excellent collection of articles on CDHPs that includes discussions of this form of adverse selection.

Three other potentially serious problems affect high-deductible policies. First, their holders may be tempted to scrimp on preventive health care measures, some of which are often among the most cost-effective. Second, a small proportion of individuals with serious chronic and acute conditions account for a large share of annual health care spending. These patients will have exceeded their maximum out-of-pocket requirements and may not have strong incentives to economize on their use of health care. Third, HSAs are more difficult to administer, and less sophisticated consumers could have difficulty distinguishing between HSAs and other options (Greene et al., 2008a). Despite these concerns, Cogan, Hubbard, and Kessler (2005) develop a well-crafted defense of HSAs along with other market-based reforms, and Cannon (2008) makes a strong conceptual case for "large HSAs" with the full amount of employer and employee contributions put into an HSA.[3]

Other Market Reforms

Other reforms are important to proponents of market-based solutions. The first deals with the tax subsidy of employer-provided insurance. We have already described the employee gains from the tax-free income associated with such insurance, and the bias it creates toward deep coverage and associated increases in utilization. Because proposals to eliminate the tax preference would meet considerable opposition, others have argued for full deductibility from taxable income of individual expenditures on health care and health insurance (Cogan, Hubbard, and Kessler, 2005).

Another reform under the competitive approach would eliminate many mandated benefits to increase the availability of lower-priced insurance policies. As of 2016, the 50 states plus the District of Columbia have almost 1,100 mandates to provides specified benefits. Many are quite common (mammogram and prostate screening, alcohol and smoking cessation), but others are less common (wigs in Rhode Island, bone mass measurement in Maryland).

Mandates raise insurance premiums and reduce the options available to consumers, especially low-cost policies. To get around the costs imposed by mandates and other state regulations, a competitive strategy would allow individuals to purchase insurance across state lines, now generally prohibited. Parente and Feldman (2008) estimate the reductions in the number of persons without insurance in three scenarios: a national insurance market (which would have the greatest impact), one with competition among states grouped into four regions, and one with competition among the five largest states. Their "moderate" projection (prior to the ACA) for a national market indicated an increase in the number of insured of more than 12 million, if Congress were to remove the interstate insurance barrier.

Graphical Representation of the Competitive Approach

We illustrate the essence of the competitive approach with the help of Figure 22.5. Let S_1 and D_1 represent the existing demand and supply curves for health care. Equilibrium quantity is Q_1, and total spending is rectangle $0P_1EQ_1$.

Competitive strategies have two broad goals. The first seeks to reduce demand by increasing the number of patients in HSAs and other settings who are sensitive to price and the diminishing marginal benefit associated with health care. Neutralizing the tax subsidy for employer-provided health insurance would decrease demand for health services, especially the relatively less-beneficial services to D_2. The equilibrium quantity will decrease, as will prices and expenditures, though the price effect will be small where the elasticity of supply is large.

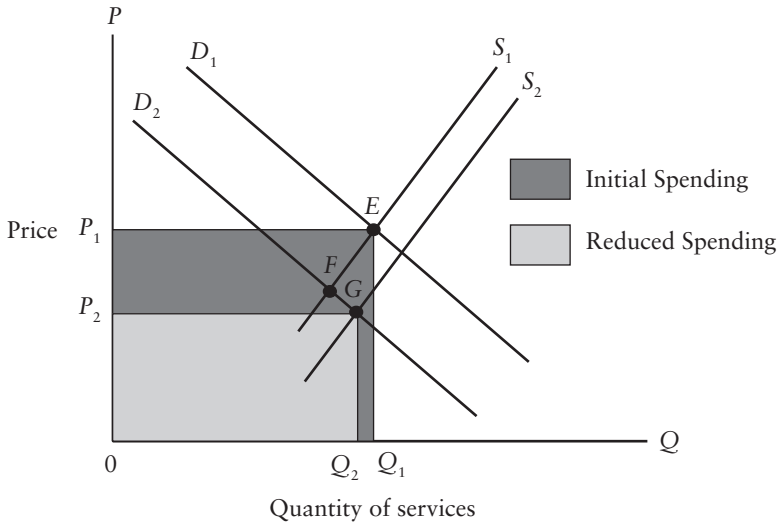


Figure 22.5 The Intended Effects of Competitive Strategies on Demand and Supply

However, another important role for competitive strategies operates through effects on the supply side. Here, advocates believe that a relaxation of regulatory, entry, and capacity controls will reduce producer costs and increase the supply of services. At the same time, competitive pressure introduced through consumer search will push providers to produce care more efficiently, that is, at lower cost, represented by a rightward shift in supply to S_2. The combined effects, if substantial, would lead to large decreases (from $0P_1EQ_1$ to $0P_2GQ_2$) in health care spending, as illustrated in Figure 22.5, or to reduced growth rates in spending. Of course, in the absence of these predicted effects, competitive strategies would be ineffective and they might even backfire.

ACA Outcomes after Six Years

There are multiple ways to measure the impact of the ACA. We will examine access, coverage, costs, quality, and adverse selection, recognizing that the implementation in each state has differed. Since many critics feared loss of jobs due to increased employer costs, we will also look at findings on employment effects. At the time of this writing (mid-2016) many findings are preliminary, but may indicate future trends.

Health Care Access

The uninsured portion of the population has tumbled since 2010. Figure 22.6 shows how the uninsured rate fell for all age groups from 2013 to 2014 (the first full year of the individual mandate and the expanded Medicaid) from 18.8 percent (35.6 million adults) to 14.4 percent (27.4 million adults). Not surprisingly the uninsured rate was higher in those states (like Texas and Florida) that did not expand Medicaid, but even in these states the percentage

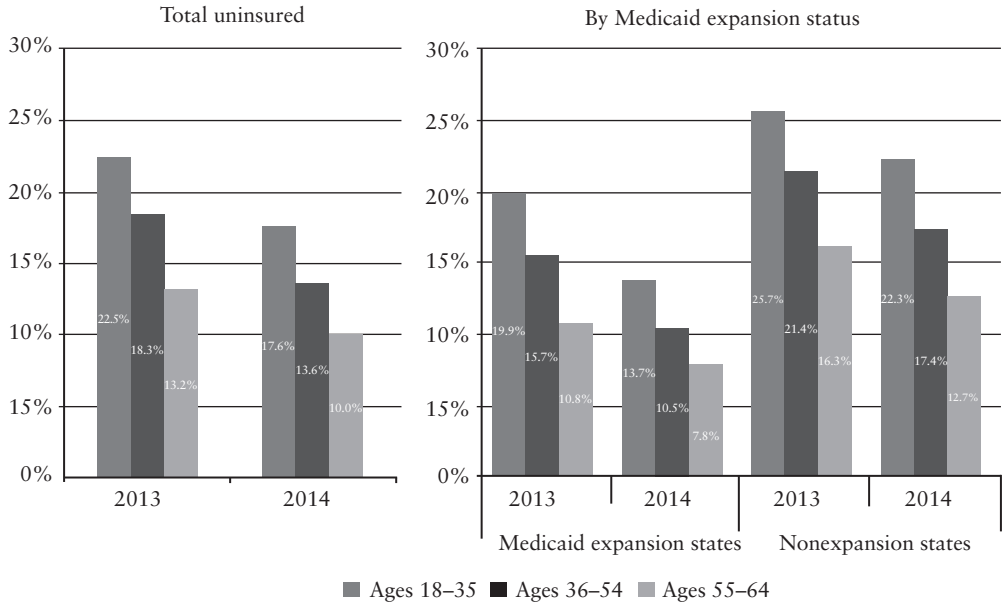


Figure 22.6 Percentage of Nonelderly Adults Uninsured for Entire Calendar Year

Source: National Center for Health Statistics, *Health Insurance Coverage: Early Release of Quarterly Estimates from the National Health Interview Survey*, January 2010–March 2015, August 12, 2005. Available at: www.cdc.gov/nchs/data/nhis/earlyrelease/Quarterly_estimates_2010_2015_Q11.pdf, accessed November 2016.

uninsured fell among all age groups. For example, the uninsured rate for those ages 18 to 35 fell from 19.9 percent to 13.7 percent in the "expansion states" compared to a fall from 25.7 percent to 22.3 percent in the nonexpansion states.

The ACA is all about health insurance. Frean, Gruber, and Sommers (2016) assess the relative contributions to insurance changes of various ACA provisions in the law's first full year, using rating-area level premium data for all 50 states and microdata from the 2012–2014 American Community Survey. Their study found only a moderate consumer response to the ACA's price subsidies. Nevertheless, the subsidies led to a coverage expansion of about 1 percent of the population (about three million people). The authors also found little impact on consumers' coverage decisions of the various exemptions and tax penalties under the ACA, and that changes in Medicaid were responsible for increased coverage of both newly eligible populations as well as those who were previously eligible but had not previously applied for Medicaid.

The "metal plans" (referenced above) appear to have increased enrollment. As of April 2016, the metal plans insured over 11 million American residents. Over 90 percent were in either silver or bronze plans (consumers' eligibility for cost-sharing reductions required enrollment in a silver plan or better), with fairly large deductibles and copayments. While the biggest numbers of enrollees were in the biggest states (California, Texas, and Florida), the metal plans had their biggest percentage impacts in states that had not expanded their Medicaid plans in response to the ACA (see Table 22.2). In terms of percentage

Table 22.2 Purchase of the "Metal" Plans—States with Highest Percentages

State	*Percent Population with "Metal" Plans*	*Medicaid Expansion?*
Florida	7.56	No
Idaho	5.70	No
Maine	5.66	No
Utah	5.49	No
North Carolina	5.43	No
Montana	5.01	**Yes**
Georgia	4.68	No
Virginia	4.52	No
Vermont	4.45	**Yes**
Nebraska	4.23	No
South Carolina	4.18	No
Missouri	4.14	No
Texas	3.98	No
Louisiana	3.95	**Yes**
Wisconsin	3.88	No

Source: Tabulations derived from "Marketplace Enrollment by Metal Level," http://kff.org/health-reform/state-indicator/marketplace-enrollment-by-metal-level/, accessed August 18, 2016.

covered in the metal plans, all of the top five states (Florida, Idaho, Maine, Utah, and North Carolina) and 12 of the top 15 percentage metal states were those that had not expanded Medicaid.

BOX 22.2

Has the ACA Improved Access to Care?

We have seen how the ranks of the uninsured diminished substantially following the ACA, but insurance does not necessarily mean easy access. Although insurance is an important determinant of access, coverage gaps or large copayments in some insurance plans, lack of transportation or child care, unavailability of providers or those who accept the patient's insurance, and discrimination may explain the wide disparities in access that have characterized U.S. health care delivery. Improved access for some population groups remains a major policy priority.

An ongoing project involving one of this volume's co-authors (Stano) provides unique information about access to care in the form of availability of a primary care physician for new patients and their wait-to-appointment. The researchers use simulated consumers (also known as audits) to examine the effects of insurance, race/ethnicity, and sex in the search for a new physician. By experimentally controlling for patient characteristics, audit studies avoid the confounding differences in patient characteristics, e.g., health status, that plague surveys or other forms of data collection. This method has been used in several forms of research that investigate economic discrimination, for example, differential behavior of housing rental agents toward whites and people of color.

In this study, research assistants (RAs) called the offices of a large random sample of primary care physicians listed in the American Medical Association's Masterfile in 2013, just prior to the individual mandate, as well as in the two following years. The RAs purportedly called on behalf of an aunt or uncle who was in good health and was either 47 years old (for those with Medicaid, private insurance, or self-pay) or 67 years old (for those with traditional Medicare). The RAs used names that signal race/ethnicity, e.g., Tamisha Washington, Juan Martinez-Hernandez, and Greg O'Brien.

For the 2013 baseline year, Sharma, Mitra, and Stano (2015) showed, for example, that the probability of an appointment offer for "Black female patients" on Medicaid was 20 percent compared to 70 percent for "White male self-pay patients." Statistical methods indicated up to five-fold differences in offers of appointment to different groups based on insurance, race/ethnicity, and sex, as well as large differences in wait-to-appointment.

In another report Tinkler and colleagues (2016) examined the availability of nurse practitioners (NPs) in 2013 if a primary care physician was not available. NPs constitute a large profession with 136,000 employed in 2015 and about one-half working in physician practices. The authors found that NPs were offered appointments less than 1 percent of the time despite their much lower wait-to-appointment times for NPs—3.6 days compared to 22.5 days for physicians.

Did access for Medicaid patients improve in the following two years? Did the newly insured patients under the ACA, as some feared might happen, make it harder for other groups to access physicians? Preliminary analysis comparing 2013 with the post-mandate years 2014 and 2015 provides some answers. Sharma et al. (2016) found that appointment offers were higher in 2015 than in 2013 for Medicare and privately-insured patients. Appointment offers for Medicaid patients increased each year with stronger effects in states that expanded Medicaid.

Nevertheless, there remained a large and persistent disparity between Medicaid and other insurance groups. For self-pay patients, there was an overall decrease in appointment offers across states that expanded Medicaid, suggesting the possibility of some crowd-out, but an overall increase across nonexpanding states.

Health Care Costs

During 2006 and 2007, immediately preceding the Great Recession, the National Health Expenditures (NHE) growth rate exceeded 6 percent. In 2009, the last year of the Recession, the rate dropped below 4 percent and remained there through 2013, or five years of apparently reduced growth. However, growth then accelerated to 5.3 percent in 2014 and 5.8 percent in 2015, coinciding with the expansion of the ACA. Are they related and if so how?

Roehrig (2016) argues that the increases in NHE during 2014 and 2015 resulted in part from expanded coverage under the ACA. Increased access to both public and private health insurance increased utilization over this period and drove up overall spending. Expanded coverage also had impacts on prescription drug spending and the cost of insurance. Roehrig also reminds us that there was a large spike in prescription drug spending in 2014 resulting from the introduction of hepatitis C drugs.

Did the ACA affect this? Yes, in two ways. Yes, there was a major increase in enrollment, which grew faster than the economy. We have seen in the past that enrollment increases lead to expenditure increases, and growth in the NHE share. Yes, in the sense that some transitions to managed care increased growth. In a transition to managed care Medicaid, the net insurance cost rises as Medicaid Health Maintenance Organizations (HMOs) collect more in premiums than they pay out in benefits while government administrative costs are largely unaffected.

Schoen (2016) argues that a number of ACA reforms, particularly related to Medicare, have likely contributed to the slowdown in health care spending growth by tightening provider payment rates and introducing incentives to reduce excess costs. Among these are various provisions that reduced payments to hospitals, other providers, and private Medicare Advantage plans; and, as described below, several incentive programs designed to improve quality and lower costs.

The ACA and Quality

The ACA established numerous mechanisms to address quality. The essential benefits requirement that insurance cover "preventive and wellness services and chronic disease management" (Table 22.1) is a clear and obvious example. The Act also contains more specific provisions that include financial incentives for quality improvement. For example, three programs authorize Medicare to link payments for hospital care to quality: The Hospital Readmissions Program (HRP); the Hospital Inpatient Value-Based Purchasing Program (VBP); and the Hospital Acquired Condition (HAC) Reduction Program. The HRP reduces payments to acute care hospitals that have excess readmission rates for certain high-cost or high-volume procedures.

The Hospital VBP program is also designed to promote quality by first reducing DRG payments to participating hospitals (about 3,000) and then using those amounts to fund incentive payments to hospitals based on performance measures relative to other hospitals or based on improvements in a hospital's performance relative to a baseline period. HAC reduces payments to hospitals that rank in the bottom quartile of hospitals based on HAC performance measures.

Medicare has also implemented many other initiatives at the same time as quality reporting, so it can be extremely challenging to evaluate the impact of any specific program or set of programs on health outcomes. Nevertheless, important and encouraging evidence has emerged on the benefits of the ACA despite the challenges and the limited timeframe since the onset of its major provisions. Jacobs, Duchovny, and Lipton (2016) used national survey data to compare self-reported health status and health care use in 2014 with 2013, just prior to the individual mandate for three population groups: Medicaid, individual nongroup, and the uninsured. The study is revealing about possible population shifts among the three groups. It found improved health status among the uninsured in 2014 compared to 2013 suggesting that those who had taken up insurance in 2014 were less healthy than the overall uninsured population in 2013. Medicaid enrollees showed improved health in 2014. With substantial Medicaid expansion under the ACA, this result

is likely due to the better health of the newly enrolled compared to existing Medicare enrollees. Those with private insurance had more chronic conditions in 2014 than in 2013. As the authors suggest, this was likely due to the newly insureds' being diagnosed with such conditions.

From a continuous national telephone survey representing a large sample of respondents, Sommers et al. (2015) examined a variety of outcomes variables covering a timeframe (2012–2015) that included the ACA's first two open enrollment periods. In addition to improvements in coverage and access following the start of the initial open enrollment period (October 2013), the study found a clear downward trend among those reporting their health conditions as "fair or poor." Analyses of subsamples, e.g., by race/ethnicity and by states that participated in Medicaid expansion vs. those that did not, revealed broad gains following the initial enrollment period.

Sommers et al. (2016) also provide a more extensive evaluation of the Medicaid expansions by conducting their own surveys of about 1,000 low-income adults in each of Kentucky, Arkansas, and Texas for the years 2013–2015. Kentucky and Arkansas expanded their Medicaid programs while Texas did not. The information collected included self-reported health status, changes in health insurance, utilization of services, and various socioeconomic measures. The study determined (p. E8) that compared to Texas, Medicaid expansions in Kentucky and Arkansas led to "major improvements in access to primary care and medications, affordability of care, utilization of preventive services, care for chronic conditions, and self-reported quality of care and health."

Employment Effects

The ACA places mandates on both labor suppliers (requiring that they buy insurance) and labor demanders (requiring that they provide insurance). We learned in Chapter 11 that these mandates can affect employment when the marginal costs and the marginal benefits of the insurance are not equal. Although the measurement of employment effects for the ACA is as yet premature (the program has not been uniformly implemented, and data must be collected on a state by state basis), Kolstad and Kowalski (2016) use a variant of our Figure 11.2 to sort out the impacts of the 2006 Massachusetts health care reform, which featured many of the same features.

Redrawn as Figure 22.7, we examine employers N that do not initially (at time b, before the reform, or Nb) offer employer-sponsored health insurance, and employers E that do (Eb). The equilibrium money wage w^{Eb} for those who offer insurance (point D) is *lower* than w^{Nb} for firms that do not (point A), because the insurance has value to the workers, who will willingly pay for it in reduced wages.

The ACA requires that most employers offer insurance or pay a penalty. Those who choose to pay will see their demand curves (at time a, after the reform) shift downward by the amount of the penalty, leading to an equilibrium reduction in money wage to w^{Na}, and a reduction in employment (from L^{Nb} to L^{Na}). Why? They are paying more, so they are hiring fewer workers.

The employers who already offer insurance see the supply curve for their workers falling. Why? Even if workers do not value the insurance on its own merits, in reacting to the individual mandate, they will value it at least as much as the penalty they must pay for not having it. Observers will see a reduction in money wage to w^{Ea}, and an increase in employment because the workers cost less.

The impact on overall employment depends on the relative magnitudes of the changes in the two sectors. Employment before the reform was $L^{Eb} + L^{Nb}$; it is now $L^{Ea} + L^{Nb}$. The net

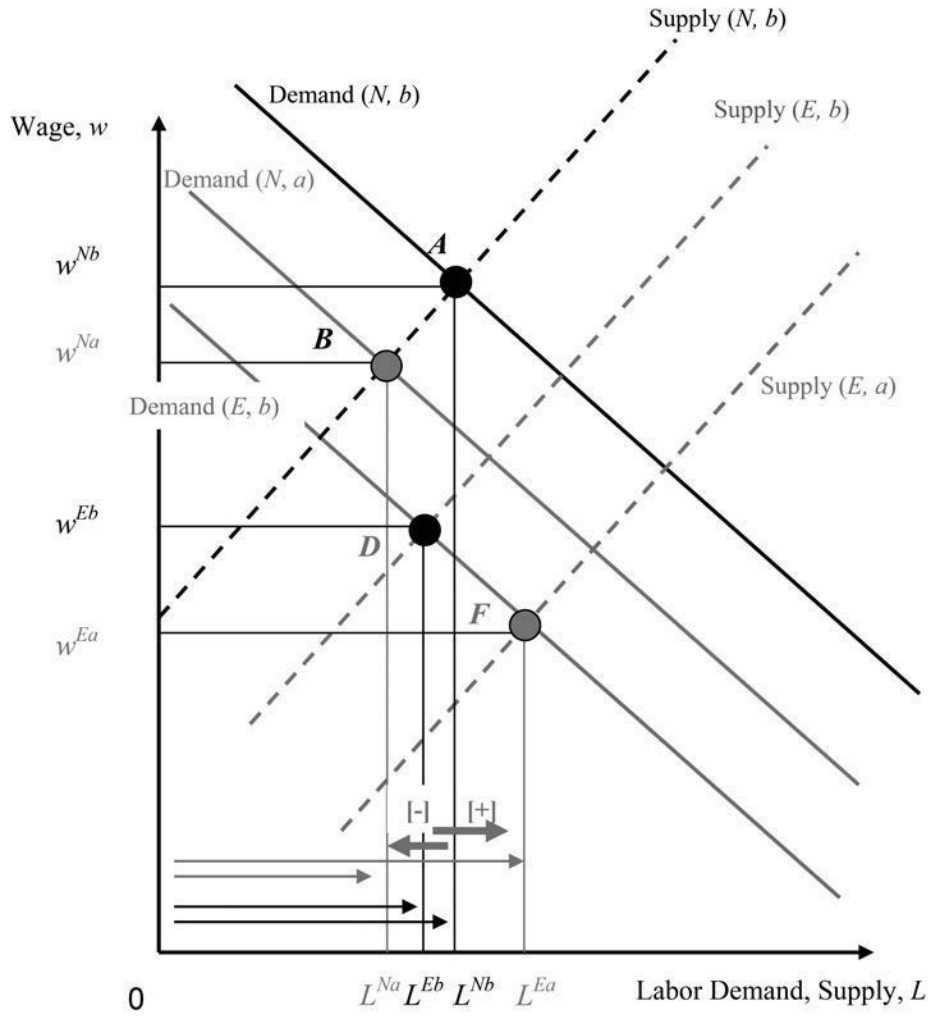


Figure 22.7 Employment Effects of Mandated Insurance through Health Care Reform

effects depend on the impacts of the mandated penalties on the employers, and the mandated coverage for the employees.

The authors report several key findings:

- Those who gained employer-supplied health insurance were willing to accept lower wages because they valued the coverage that they received.
- The Massachusetts mandating mechanisms were desirable from a welfare economics perspective. The authors find that if the government had established a wage tax to pay for health insurance, the economic losses due to that labor market distortion would have been about 13 times as large as they measured.
- Finally, contrary to the hypothesis that employers would simply stop offering coverage, employer-supplied coverage increased. The authors believe that the individual mandate,

combined with the valuation of employer-supplied coverage, encouraged workers to demand employer-supplied coverage from their employers. This was paid for out of decreased wages. If employers have a comparative advantage in offering health plans that their employees value (due to increased risk experience and more risk pooling), one might expect a "crowd-in" to employer-supplied coverage, precisely what the authors saw.

Adverse Selection under the ACA

Adverse selection is a phenomenon that we explored in Chapter 10. It can arise in insurance markets when some buyers are able to purchase insurance below actuarially fair rates. The ACA imposed an individual mandate but, aside from their age and tobacco use, higher-risk consumers or those with pre-existing conditions cannot be denied coverage or charged a higher premium. Have these features created opportunities for adverse selection?

The *New York Times* reported that those enrolling in Blue Cross plans in 2014 and 2015 were sicker with higher rates of "diabetes, depression, coronary artery disease, H.I.V. and hepatitis" than those with prior coverage. Hospital admissions were 84 percent higher and visits to physicians and other providers were 26 percent higher.[4] Another *New York Times* article described the numerous waivers granted by the Obama administration that allowed "special enrollment" periods beyond the annual enrollment deadline for certain groups. Those who enrolled during these special enrollment periods had 55 percent higher utilization than those who enrolled before the regular deadline.

Adverse selection by potential enrollees is not the only problem. Many insurance plans are finding that large numbers of new enrollees drop their coverage soon after they have signed up. According to the *Detroit Free Press*, Michigan Blue Cross has reported losing an unexpectedly high 20 percent of its ACA subscribers in the months following open enrollment. The enrollees get needed services and then drop their coverage, presumably paying the tax penalty for the remaining months that they do not have coverage.

We caution that scholars have not yet thoroughly analyzed the ACA evidence. Hackman, Kolstad, and Kowalski (2015) examined the individual mandate under the Massachusetts reform model, prior to the enactment of the ACA. Massachusetts had already prohibited insurers from discriminating on the basis of pre-existing conditions so its experience may differ from the national experience following the ACA mandate. The Hackman study found that Massachusetts markets experienced adverse selection prior to the state's reform and that there was a reduction in adverse selection following its reforms. Premiums and costs in the individual insurance markets decreased significantly.

Following the ACA, at the national level there were wide variations in premium increases across states in 2015 and 2016 but the average was moderate (Blumberg, Holohan, and Wengle, 2016). Nevertheless, large effects reported by the media are potentially troubling. Insurers may be less likely to participate in the exchanges, reducing competition and consumer choice, and those insurers that remain may seek higher rate increases to cover their costs. As of mid-2016, California insurers requested an average premium hike of 13.2 percent for 2017, and many major insurers in other states were also requesting approval for double digit rate increases.

Meeting Reform Goals

Six years after enactment, how well has the ACA addressed the reform goals discussed earlier in this chapter?

Creating a Safety Net

Millions of American residents still do not have health insurance and many who do face substantial cost sharing in the forms of deductibles and copayments. While analysts accurately foresaw that the program would not insure everyone, they did not foresee that two of the three largest states (Texas and Florida) would refuse to expand the Medicaid programs, leaving several million people without coverage. All that said, however, from January 2014 through 2015, the percentage uninsured dropped rapidly from 15 percent to less than 9 percent, thus becoming the lowest ever recorded in the United States.

Cost Containment

At the beginning of the 2010s, the U.S. health economy entered a period of reduced cost growth (which is *not* the same as reduced costs). As noted above policies that accompanied the ACA may have reduced cost pressures, although it is difficult to point to explicit ACA policies that did so. However, cost growth began to accelerate in 2014 and 2015, so analysts will need longer-term data before they can explicitly link cost changes to ACA policies.

Quality, High-Value Care

Improving health care quality is challenging to any health care system, but early evidence of the impacts of the ACA's Medicaid expansion has been encouraging. Several Medicare initiatives for financial rewards to providers that improve quality and lower costs also show considerable promise.

Choice for Patients and Providers

The decision to use the existing network of providers, in a framework to preserve the private insurance system, has left much of the patient choice unchanged. In general, patients have no less provider choice than they had prior to the passage of the ACA.

Ease in Administration

The state-by-state implementation of the ACA has not led to ease in administration, compared, for example, with the federally administered Medicare program, or a single-payer Canadian-type of system.

Conclusions

Cost containment, and reduction or elimination of the number of uninsured, are the principal goals of health system reform in the United States. Other goals include administrative simplicity and choice for providers and patients. Improving the quality of care has also emerged as a national priority. Any reform process requires difficult decisions on the services covered, on who is covered, and on the financing mechanisms.

The most serious obstacle to reform (using the United States as our example) has been the divide over whether to expand the government's role through mandates, additional regulations, and tax subsidies, or whether to rely increasingly on markets through deregulation and tax changes that neutralize the current bias toward subsidized, employer-based insurance.

We have examined other important issues to health system reform. We found that the incidence of health premiums under employer-based systems falls on workers. In theory, the burden of increases in health care costs or mandates on employers will fall on workers. Thus employer mandates do not make firms less competitive internationally, nor will movements toward single-payer systems funded by government revenues make them more competitive.

The ACA passed in 2010, and after two favorable Supreme Court rulings, has produced substantial reform. Built on the shoulders of America's private insurance and the Medicaid/CHIP systems, it uses an individual mandate for consumers to purchase health insurance and provides market-pooling mechanisms to make the insurance available to many who were previously not able to get it. It has reduced the number of uninsured in the country by over 16 million with the gains coming primarily through the marketplaces (health exchanges), young adults' staying on their parents' plans until they turn 26, and Medicaid expansions.

Summary

1 Many difficult decisions in the reform process include determining which services to cover, who to cover, and how to pay the additional costs.
2 For society, the cost of universal coverage is the incremental cost of additional health care purchased by people due to improved insurance coverage.
3 The incremental cost of providing full-year coverage for all Americans would be about 5 percent of current NHE. Savings from successful cost controls may reduce this incremental cost.
4 Mechanisms to reduce the uninsured include employer and individual mandates, expansion of existing public programs, and subsidized coverage for lower-income and high-risk households.
5 A switch from the current U.S. health system to a single-payer system would likely save money by reducing administrative costs.
6 For the entire country, labor supply is inelastic. As a result, the incidence of a mandated health insurance program falls mainly on workers.
7 Mandates on employers have the same economic effects as mandates on workers.
8 Some proposals recommend the separation of health insurance from employment in order to eliminate the inefficiency caused by the tax-exempt status of employer-provided health insurance.
9 Competitive strategies include the promotion of alternative delivery systems, the expansion of consumer-directed health plans built around various forms of health savings accounts, reductions in mandated benefits and other regulations on the insurance industry, and implementing tax reforms that reduce the bias toward employer-based insurance.
10 The ACA seeks to reduce the number of uninsured individuals by improving the affordability of insurance, and by improving the availability of coverage for employees of small businesses, in the context of the existing U.S. health insurance industry.
11 The ACA provides two avenues for expanding insurance: (1) providing exchanges where consumers can purchase insurance in competitive markets; (2) expansion of Medicaid, by which several million Americans can get treatment through established providers.
12 In bypassing either single-payer programs or explicit mandated providers, the ACA preserves both patient and provider choice. Consumers maintain choice of insurance coverage, and provisions for provider payment are unchanged.

13 The ACA's cost-containment strategies are less well-developed than those that provide the insurance. Some overall policies related to Medicare appear to have reduced cost pressure in that program. In addition, incentives to reduce hospitalization and hospital time, as well as increased accountability, appear to have had some cost-dampening impacts.
14 Health insurance enrollment has risen, and the number of uninsured has fallen by 20 million from its pre-ACA level of 2009. Overall healthcare costs (both aggregate and percentage) have risen due to an increase in the number of enrollees and the structures of the new plans.

Discussion Questions

1 Would individual mandates for health insurance be more or less burdensome to the poor than employer mandates? Would lower-income groups be wise to favor one plan over the other?
2 If the aggregate labor supply curve were highly responsive to increased wages (elastic) instead of very inelastic as we have stated, how would an employer-mandated health insurance plan affect the country's international competitiveness?
3 How could a single-insurer health insurance system provide additional cost savings over a multiple-insurer system like that in the United States?

Price	*Quantity*
100	100
90	200
85	250
80	300
70	400
60	500
55	550
50	600
40	700
30	800
20	900
10	1,000
0	1,100

4 What are the major competitive strategies proposed by economists? How, for example, would elimination of the tax subsidy for employer-provided health insurance reduce spending on health care?
5 What are Health Savings Accounts (HSAs)? Why might HSAs decrease health care spending? What are some problems with the HSA concept? Consider an enrollee in a CDHP with a high-deductible HSA who is choosing between the two physicians whose fees vary by $50 per visit. How might the fee information influence her to choose the lower-priced

physician? Why might the patient still choose the higher-priced one? Would your answer change if the difference was $200 per visit?

6 Compare the ACA passed in 2010 with the Massachusetts health plan adopted in 2006 (use the Internet to obtain more details than we have provided in this chapter). Be sure to note the similarities and differences in the two plans.

7 Improving the quality of health care is becoming a national priority. Discuss the relative merits of using government regulation versus market forces in improving quality.

8 Assume that a brilliant health economist has developed a plan that will greatly improve the efficiency of the U.S. health care system by putting everyone's medical records on a card with a readable chip. Would it be likely that this plan, or anything close to it, would be adopted in the United States? (Hint! Think of those who might be interested in preserving the status quo.)

Exercises

1 Suppose that a monopolistic firm faces a downward-sloping demand curve for its product and offers no health insurance to its employees. Let an employer mandate for health insurance be enacted. If this causes the firm's marginal costs to increase, will the firm pay the full cost of the health insurance out of profits? Is the mandate likely to increase the firm's marginal costs? Provide a diagrammatic discussion and solution.

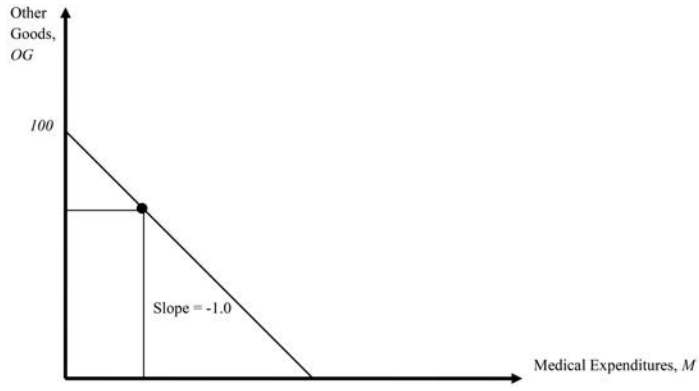
2 Suppose that a purely competitive firm offers no health insurance to its employees. Let an employer mandate for health insurance be enacted. If this causes firms' marginal costs to increase, will the firm pay the full cost of the health insurance out of profits? Is the mandate likely to increase the firm's marginal costs? Provide a diagrammatic discussion and solution for the firm and the market.

3 Suppose that a monopolist faced the following demand curve for its goods. Its marginal cost per unit of production is 50, and it faces no fixed costs
   (a) Calculate the profit-maximizing output and price.
   (b) Suppose the workers negotiate a health insurance benefit increase that increases marginal cost per unit from 50 to 60. Calculate the new profit-maximizing output and price.
   (c) Who bears the costs of the benefit increase? Why?

4 Consider a diagram like Figure 22.3 where states are offered the opportunity to buy into Medicaid under the ACA for a 9:1 match.
   (a) Mark an initial level of expenditures if the state is spending $20 on Medicaid expenditures and $80 on everything else, with an appropriate indifference curve.
   (b) Suppose the state decides to collect a tax of $3, from the initial point, to put up for the match. How big will the match be? Draw the tax and the match on the graph.
   (c) In this example is the state likely to be better off or worse off by collecting the tax to get the match? Why or why not?

5 Consider Figure 22.4 where health insurance and other goods are measured in dollars spent. Suppose Adam spends no money on health insurance, Steve spends 6 percent of his income on health insurance, Beth spends 10 percent of her income, and Jackie spends 15 percent of her income.
   (a) Draw each person's location on the budget constraint.
   (b) Suppose that a government program offered at no cost a level of insurance equal to 5 percent of one's expenditures. Indicate this on the graph. Which person(s) would be certain to "take up" the new policy? Which would be certain not to take up the policy?

(c) Redraw figure (b) with a program that was offered with a 50 percent subsidy, rather than without cost. Which person(s) would be certain to "take up" the new policy? Which would be certain not to take up the policy? Would your answers vary from answer (b)? Defend your reasoning.



6 Consider two households. They have the same incomes and face the same prices. Household *H* tends to be healthy and household *U* tends to be unhealthy. Suppose that two insurance plans are available:
A—$2,500 deductible and a 5 percent coinsurance rate after meeting the deductible.
B—$250 deductible and a 20 percent coinsurance rate after meeting the deductible.
   (a) Using a budget constraint and indifference curves on the diagram above, model the two insurance plans.
   (b) Assume that a voluntary HSA is made available upon the purchase of a high-deductible policy. Assume that if the money is not used it is lost. Which of the households is likely to participate? Use the diagram above to explain why.
   (c) Consider part (b) above, but assume now that the unused portion in the HSA can be distributed to the individual at the end of a designated period or at retirement. Would your answer to part (b) change? If so, how? If not, why not?

7 Use the demand–supply framework in Figure 22.5 to explain how increased cost sharing could lead to lower utilization and spending on health care.

8 Suppose that the "greasy food" sector of the economy did not offer health insurance before the passage of the ACA, and "greasy food" workers were not interested in health insurance.
   (a) Draw and label an initial labor market equilibrium where the wage equaled $9/hour.
   (b) Assume that after passage of the ACA, the "greasy food" owners chose to pay a $0.50 per hour fine rather than offer health insurance. What would happen to the market wage in the sector? What would happen to employment in the sector? Explain your answer.

9 (Advanced) Individual or class project. For students with computing and statistical skills, the MEPS database is available at www.meps.ahrq.gov/mepsweb/. For an individual or class project, try to replicate, or improve, Coughlan's 2014 estimates of the costs of universal insurance for the most recent data year. Be attentive to key assumptions as to how much care those who are currently uninsured might purchase, under the various proposals.

Notes

1 The Healthcare Effectiveness Data and Information Set (HEDIS) is a tool, developed by the National Committee on Quality Assurance, used by more than 90 percent of America's health plans to measure performance on important dimensions of care and service.

2 *Republican Platform 2016*, p. 36. Available at https://prod-static-ngop-pbl.s3.amazonaws.com/media/documents/DRAFT_12_FINAL[1]-ben_1468872234.pdf, accessed August 4, 2016.

3 Other countries have introduced similar accounts called Medical Savings Accounts or MSAs. Singapore introduced MSAs in 1984 and several other nations, most notably China, have adopted MSA options. Hurley et al. (2008) simulate the effects of a publicly funded MSA system for Ontario, the most populous Canadian province. Their simulations indicate some cost savings but also adverse distributional effects on public spending and out-of-pocket costs.

4 Sources for this section include Robert Pear, "Newest Policyholders under Health Law are Sicker and Costlier to Insurers," *New York Times*, March 30, 2016: http://nyti.ms/2301aLr, accessed November 2016; Robert Pear, "Insurers Say Costs Are Climbing as More Enroll Past Health Act Deadline," *New York Times*, January 10, 2016: http//nyti.ms/1OYMuCM, accessed November 2016; J.V. Reindl, "Affordable Care Act Rates may Jump by 17.3% in Michigan," July 31, 2016: www.freep.com/story/money/business/michigan/2016/07/30/obamacare-rates-affordable-care-act/87623260/, accessed November 2016; Claudia Buck, "Covered California Health Care Premiums to Jump 13.2% in 2017," *The Sacramento Bee*, July 19, 2016: www.sacbee.com/news/local/health-and-medicine/article90542787.html, accessed November 2016.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 23

The Health Economics of Bads

In this chapter

- An Introduction to Bads
- Models of Addiction
- Rationales for Public Intervention
- Advertising Restrictions on Cigarettes and Alcohol
- Excise Taxes and Consumption of Cigarettes and Alcohol
- Conclusions

No discussion of the health economy can be complete without addressing one of the most prominent and controversial issues—the consumption of health bads. In a market economy, the consumer's sovereignty usually is regarded as given, that is, the consumer is "free to choose." However, we can find many exceptions in any society—cases where society encourages or discourages desired choices. We encourage and promote use of motorcycle helmets, automobile emissions-control devices, old-age pensions, and good prenatal care. We often discourage the purchase of alcohol, street drugs, and cigarettes.

The reason to intervene in private decisions could be paternalistic, but it also could appeal to economic efficiency. Cigarette smoking affects not merely the cigarette buyer and seller—effects that are internal to the cigarette market—but also the health of nonsmokers nearby, an external cost. Since many health insurers do not distinguish between smokers and nonsmokers, perhaps because of high monitoring costs, nonsmokers may pay higher premiums than warranted by their actual health risks. The personal hazards of excessive alcohol consumption are also serious, including disability due to alcoholism and fatality due to liver disease. The external costs include possible harm to family or neighbors as well as the excessive, dangerous, and often-fatal traffic accidents due to drunken driving. Finally, when smokers and drinkers underestimate the probabilities of ill health due to their consumption, the imperfect information provides an efficiency rationale for measures, such as taxes, that tend to curb the behaviors.

Economists as citizens may choose one side or the other of these controversies because their values are not determined by their being economists, but economists as scientists can illuminate the issues that are of interest to the public. For many, the question is not whether to intervene in private decisions to smoke or drink, but how to do so more effectively and unobtrusively. This hinges on many issues although two economic questions are central. First, what is the relationship of price to demand? Second, what is the relationship of product advertising to total consumption?

An Introduction to Bads

Economic reasoning is helpful for any approach to the analysis of bads, such as cigarette smoking and excessive alcohol consumption. Box 23.1 shows how consumption of alcohol and tobacco varies across countries. Economic models of addiction as well as market failure help determine whether intervention is justified on efficiency grounds. Alternatively, models of consumption, advertising, price, and taxation help provide relatively unobtrusive and cost-effective means to intervene, once intervention is chosen.

BOX 23.1

Who Smokes and Who Drinks? Cultures and Behaviors

It is common enough to find families who drink wine with dinner, or groups of people who smoke together. A comparison of the consumption across countries gives some perspective. It may surprise some Americans that U.S. consumption per capita of both "bads" is relatively low among the developed countries shown.

In Martin Cruz-Smith's popular novel, *Gorky Park*, the Russian central character, Arkady, is asked, "Why didn't you ever go to America?" Arkady answers, "If I went

to America, I would have had to quit smoking." Alcohol consumption data and cigarette smoking prevalence data are from *OECD Health at a Glance 2015*. Comparable Russian data are not available.

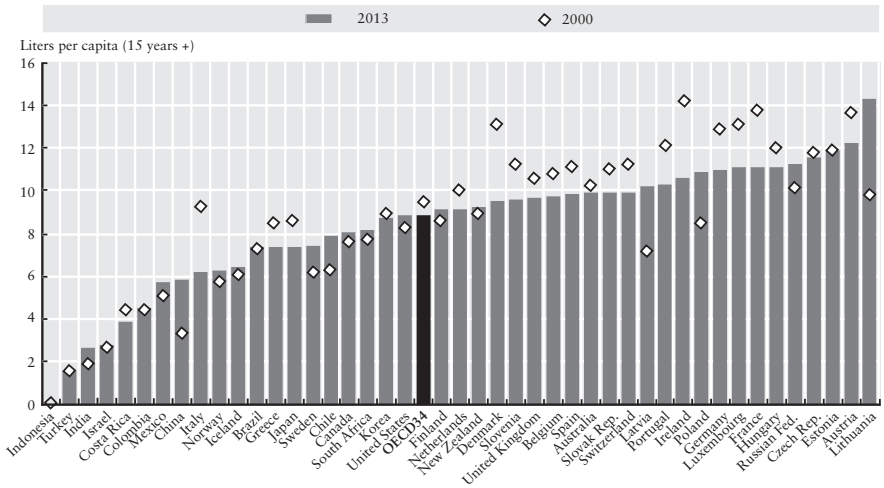


Figure 23.1 Alcohol Consumption Age 15 and over, 2000 and 2013

Source: OECD (2015).

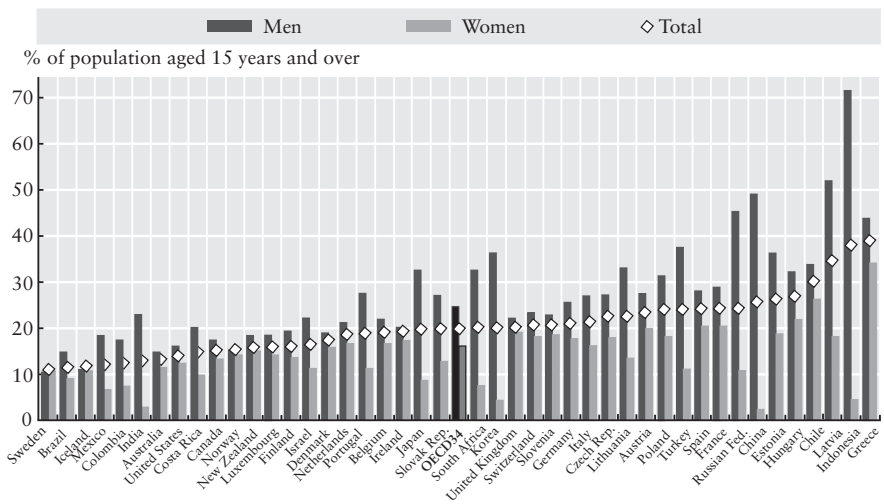


Figure 23.2 Daily Smoking in Adults, 2013 (or nearest year)

Source: OECD (2015).

That cigarettes qualify as bads follows from the high death rates attributed to cigarette smoking. Deaths per capita from respiratory cancers rank highest among national death rates from malignancies and second only to categories of heart disease overall. Motor vehicle deaths are less than half this rate, and only a handful of states have HIV/AIDS death rates that exceed one-half of the respiratory cancer rate. Cigarette smoking is known to affect other disease categories as well, such as emphysema and heart disease. There is little doubt that getting successive generations of youth to choose never to smoke would lower average population age-adjusted mortality rates substantially. Econometric studies of health production commonly find cigarette consumption to be a significant and materially important predictor of mortality rates.[1] Table 23.1 confirms these comparisons of mortality rates.

In contrast to smoking, moderate alcohol consumption does not harm many groups of people and is reported to benefit some. Yet, substantial harm occurs with excessive consumption and inappropriate related behaviors, such as drunk driving. Applying a novel means to identify the effect of drinking on traffic fatalities, Levitt and Porter (2001) found that drivers with "alcohol in the blood" are eight times as likely to cause a fatal accident as the sober driver. With "blood alcohol above legally drunk," the ratio rises to 15 times. Alcohol use among both high school and college students also raises policy concerns. College students who drink have poorer academic attainment (Cook and Moore, 1993; Williams, Powell, and Wechsler, 2003). High school drinkers are more likely to drop out before graduation (Chatterji and DeSimone, 2005).

One economic approach is to assume that there are no grounds to intervene if the consumer chooses rationally and voluntarily, is reasonably informed of the risks, and creates no side effects for others. This view implicitly rejects all arguments not based on economic efficiency, including those grouped together as equity concerns. Furthermore, the view typically rejects other grounds as paternalistic. However, because alcohol and cigarette consumption are addictive, the issues of rationality, volition, and information take on closer scrutiny.

Table 23.1 U.S. Mortality Rates—Selected Causes

Cause of death	*Number*	*Percent of total deaths*	*Mortality rate per 100,000*[1]
All causes (2014)[2]	2,626,418	100.0	823.57
Diseases of heart	614,348	23.4	192.64
Cancer	591,699	22.5	185.54
Accidents	136,053	5.2	42.66
Cerebrovascular diseases	133,103	5.1	41.74
Alzheimer's disease	93,541	3.6	29.33
Diabetes mellitus	76,488	2.9	23.98
Suicide	42,773	1.6	13.41
Smoking-attributable mortality[3] (2010–2014 annual average)	**600,400**	**22.9**	**188.27**

Table 23.1 *continued*

Cause of death	*Number*	*Percent of total deaths*	*Mortality rate per 100,000*[1]
Alcohol deaths (2006–2010 annual average)[4]	**88,000**	**3.4**	**27.59**
Drug overdose deaths (2014)[5]			
Prescription drugs	25,760	1.0	8.08
Opioid analgesics	18,893	0.7	5.92
Benzodiazepines	7,945	0.3	2.49
Illicit drugs	17,465	0.7	5.48
Cocaine	5,415	0.2	1.70
Heroin	10,574	0.4	3.32
Marijuana (2013)[6]	0	0.0	0.00

Notes: [1] 2014 Population: 318,907,401 (U.S. Census Bureau), http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk, accessed August 2016.

[2] CDC, Deaths: Leading Causes for 2014, National Vital Statistics Reports, Vol. 65, No. 5, June 30, 2016, p. 17, www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_05.pdf, accessed August 2016.

[3] Average annual mortality 2010–2014, includes deaths attributable to smoking from cancers, cardiovascular diseases, pulmonary diseases, metabolic diseases, perinatal conditions, residential fires; as well as lung cancer and coronary heart disease attributable to second-hand smoke. National Center for Chronic Disease Prevention and Health Promotion (U.S.) Office on Smoking and Health. Atlanta (GA): Centers for Disease Control and Prevention (U.S.), 2014. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General, Chapter 12, Table 12.15, www.ncbi.nlm.nih.gov/books/NBK294316/#ch12.s26, accessed August 2016.

[4] CDC, www.cdc.gov/features/alcohol-deaths/index.html, accessed August 2016.

[5] Includes deaths with underlying causes of unintentional drug poisoning, suicide drug poisoning, homicide drug poisoning, or drug poisoning of undetermined intent. NIH, Overdose Death Rates, www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates, accessed August 2016.

[6] Drug War Facts, www.drugwarfacts.org/cms/Causes_of_Death#sthash.XNGFHVsR.dpbs, accessed August 2016.

Models of Addiction

Models of addiction come out of both psychology/medicine and economics. Chaloupka and Warner (1999) identify three types.

Imperfectly Rational Addiction Models

These models propose that the addict has stable but inconsistent preferences in the short run as opposed to the long run. As Schelling (1978) described this person:

> Everybody behaves like two people, one who wants clean lungs and long life and another who adores tobacco. . . . The two are in a continual contest for control.
>
> (p. 290)

Are people really like this? Regret, in fact, is fairly common. In film, there is even the cliché of the hero in comedy who enters the tiger's cage telling the bystanders to ignore him should

he change his mind and even scream for help. Of course, he changes his mind. Is he two different people at those two moments?

Myopic Addiction Models

Nearsightedness about the future harmful effects of the ingested drug provides a variant of the imperfectly rational model. Here, the individuals don't see the facts clearly; they are naïve about the nature of the drug and its side effects. One may imagine someone easily persuaded by cigarette-smoking friends who may denigrate the societal information about cigarettes. One may see the teenager who only dimly perceives the realities of distant future health events, including cancers, and perhaps fails to connect the experience of smokers in older generations with his or her own behaviors and prospects.

Rational Addiction

Can addiction in some cases be a rational choice? Becker and Murphy (1988) discarded myopic models and investigated addiction by assuming that people incorporate rationally all information, past, present, and future, into their utility calculations. They showed that rationally choosing to ingest an addictive drug was possible under restrictive yet plausible conditions. The discussion here is developed from work by Becker and Murphy (1988); Becker, Grossman, and Murphy (1991); and MacDonald (2004).

Addiction researchers usually speak of "reinforcement" and "tolerance."

Reinforcement means that greater past consumption of addictive goods, such as drugs or cigarettes, increases the desire for present consumption. In short, smoking more now may make us smoke more later.

Tolerance occurs if the utility from a given amount of consumption is lower when past consumption is greater. This suggests that the future impacts of smoking or drinking or ingesting drugs decrease, when we consume more now. A single glass of wine may make someone tipsy the first time he or she drinks. Over time, with more drinking experience, the first glass of the evening may have little or no impact. We will use smoking to illustrate important model relationships, although drinking, illicit drugs, or even common substances such as caffeine, can provide similar examples.

Becker, Grossman, and Murphy introduce the construct of "addictive capital stock," S. For example, with more smoking experience, the smoker's attitude toward cigarette consumption is likely to change. We assume therefore that addictive stock "reinforces" cigarette consumption, C, meaning that the more stock, the more one will smoke, leading to curves A^1 and A^2 in Figure 23.3. Though not shown in the figure, the utility function shows the smoker as gaining utility from cigarette consumption, C, from the addictive stock, S, as well as from income, which allows the purchase of other goods in addition to cigarettes.

The important questions in the model deal with what happens over time. For example, current consumption increases the addictive stock. Listening to Mozart at age 21 will likely increase our enjoyment of Mozart at age 22, thus increasing our musical "capital stock." Most smokers will remember how bad the first cigarette tasted, but similarly smoking or drinking at age 21 may increase enjoyment of smoking or drinking in subsequent years. So, a larger addictive stock makes future consumption more pleasurable.

A myopic, or nearsighted, addict looks solely at the reinforcement effect. A rational addict, however, also considers the future harmful consequences of current addictive behavior. The rationally addicted smoker weighs the present pleasure against both the future health consequences and the beneficial impact of current consumption on future consumption enjoyment.

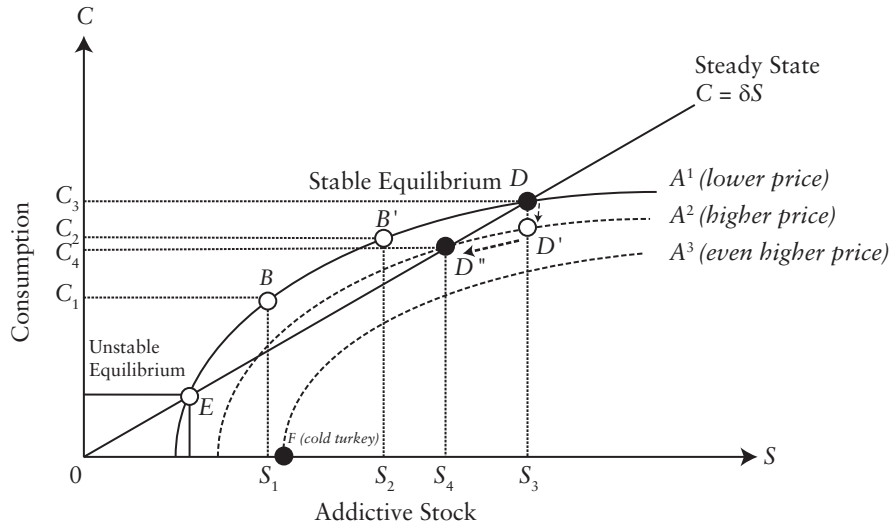


Figure 23.3 Rational Addiction Model

The rational addiction theorists have drawn several further implications from their analyses.

- Addiction is more likely for people who discount the future heavily, because they pay less attention to the potential adverse consequences.
- Addiction is more likely when the effects of past consumption depreciate less rapidly.
- Expected rises in future prices will have a dampening effect on current consumption, much like increases in current prices.

Models that examine people's behavior over time typically search for a "steady state," an equilibrium where all "outflows" and "inflows" maintain the system, like the equilibrium of a well-run fish tank. In the steady state equilibrium proposed here, the system will be maintained over time provided that current cigarette smoking adds exactly enough C to the addictive stock to replace the depreciation δS of that stock during the period. Mathematically, $C = \delta S$, where δ is the constant depreciation rate. The $C = \delta S$ line in Figure 23.3 depicts all the combinations of C and S that yield a steady state equilibrium.

Reviewing the elements of the model, we see:

1 ***Consumption of cigarettes as a function of addictive stock.*** Curve A^1 relates smoking to addictive stock for a person with a given rate of time preference (relating future utility to present utility) and a given level of wealth, and who faces a set of prices for cigarettes and nonaddictive goods. We can think of curve A^1 as a cigarette consumption curve, so the more stock S, the more consumption C. In other words, any given stock S is just sufficient to maintain consumption level C.
2 ***Cigarette consumption needed to maintain addictive stock.*** The stock of addictive capital depreciates at a rate of δ (between 0 and 1) per year. Depreciated stock is replaced with more smoking. The ray from the origin, $C = \delta S$, is the *steady state line* where current cigarette consumption just offsets the depreciation of the stock of smoking capital.

The model provides a convenient way to see what happens to the rational addict over time. Consider a smoker who is on consumption curve A^1 (as an exercise, explain why a price increase from "low" to "high" would shift the consumption curve from A^1 to A^2) with addictive stock S_1. This stock implies a cigarette consumption of C_1, or point B. Notice, however, that consumption C_1 will more than replace the depreciation in S_1 during the period (point B lies above the steady state line, $C = \delta S$). It follows that addictive stock will grow and exceed S_1 in the next period; for example it may rise to S_2 at point B'.

Continuing this logic, it follows that whenever consumption, C, lies above the steady state line, the addictive stock, S, will grow. Finally, stock S_3 and consumption C_3 give a steady state equilibrium for case A^1. We label this equilibrium point as D. Compare steady state equilibrium D with another place where the two curves cross, point E. Notice that D is like a magnet; any stock near S_3 is pulled toward D. That is, any stock a bit to the left of S_3 will bring growth in stock up to S_3; any stock to the right will depreciate down to S_3. Point D represents a "stable equilibrium." Try the same experiment with the equilibrium at point E, and see that it is unstable. Any stock slightly to the left of E will be pulled farther to the left; any stock slightly to the right will eventually increase all the way to S_3.[2]

This model provides important policy implications about impacts of price changes, often induced through tax policies. Starting at point D consider a rise in price so that the cigarette consumption curve shifts from A^1 to A^2. With the price rise, smoking decreases at first from C_3 (at point D) to point D'. It then falls farther over time since D' is below the steady state line. Equilibrium smoking level falls to C_4 at point D''. This shows the model to be consistent with our prior conceptions about price and quantity demanded. The higher the price, all else equal, the lower the quantity demanded. Moreover, the long-run responses to price changes *exceed* short-run responses. Initial decreases in smoking cause a subsequent decrease in the stocks of addictive capital, which then stimulate further smoking decreases.

Second, at some point, a rising price will cause all equilibria to disappear. Starting from point D'' and letting the price rise even more, a new A^3 curve will be everywhere below the $C = \delta S$ steady state line. This prediction is unique to the rational addiction model, the prediction that some cigarette smokers quit "cold turkey," without gradually reducing consumption down to zero, landing at a point similar to F where consumption equals 0.

It also follows that expectations about future prices of cigarettes will affect the addicts' current decisions about smoking. In Figure 23.3, we would interpret this by saying that a permanent price increase shifts the consumption curve downward farther than a temporary price increase. Both the price effect and the probability of going cold turkey increase with permanent price increases. Likewise permanent restrictions on the advertising of cigarettes would have more effect than temporary ones.

Rationales for Public Intervention

Are cigarette smokers well-informed about the risks of smoking so they can make rational choices? Some economists (Lundborg and Lindgren, 2004; Viscusi, 1995) have challenged the common wisdom that smokers are ill informed, reporting smokers' knowledge and responses to risk to be similar to that of nonsmokers. Despite this, findings from various fields regarding smoker and nonsmoker behavior more often suggest that smokers differ from nonsmokers on these bases.

Consider nonsmokers first. We mentioned that another justification for intervention is that smoking has detrimental side effects on others, or "external costs." Much then depends

on the size of these costs. These externalities come from two principal sources: the passive smoking incurred by people nearby, and various other external costs caused by health hazards to the smoker. Manning et al. (1991) estimated the external costs at the equivalent of $0.33 per pack for a new smoker evaluated in 1995 dollars, though passive smoking costs were omitted. Viscusi (1995) also estimated the external costs per pack to be in this range, though lower than Manning's, and near zero under some scenarios.

Because economic efficiency is only one economic criterion (the other is equity), because data issues can be disputed, and because economics is not the only basis to consider, many choose to intervene in tobacco and alcohol use. Economics offers two major tools that may be effective in curbing consumption of a targeted bad: restrictions on advertising and imposition of excise taxes. Advertising can be restricted by tax code revisions, but most often the public issue is whether to restrict advertising by selective or total bans. The excise tax tool is theoretically effective to the degree that demand is more elastic for a given supply. These two principal tools of intervention form the subject of our next investigations. First, however, consider why we have chosen not to address the several other tools that might be used.

Other Interventions

Two other potential forms of intervention are prohibitions on the consumption of the product and penalties for consumption or misuse of the product. Outright prohibition of cigarettes is unlikely to occur in the United States. Accounts of the period when alcohol became illegal in the 1920s seem to point to that clear conclusion. Furthermore, alcohol prohibition of that era probably was not even effective in reducing alcohol consumption (Miron, 1999; Dills, Jacobson, and Miron, 2005).

Lesser prohibitions, however, often have been accepted and proven effective. These include the effect of lower speed limits on rates of fatal accidents involving alcohol, as well as bans on smoking in public places.[3] Many countries, especially in Scandinavia, apply much more severe penalties for drunk driving than does the United States. These include stiff jail sentences for alcohol offenses, offenses that many Americans regard as less serious. Mullahy and Sindelar (1994) showed for U.S. data that legal penalties affect drunken-driving behavior. Increased fines and license revocations significantly reduce the individual's probability to drive drunk.

Regulation of smoking sometimes fails to work, although some research finds reductions in smoking after the passage of clean air restrictions. Tax effects might even be somewhat overestimated if part of the "tax effect" is really due to unmeasured local clean air restrictions. In a study of Canadian smokers, Lanoie and Leclair (1998) find that cigarette demand responds to taxes (elasticity of –0.28) but not to regulation, while the converse is true for the proportion of smokers in the population.

Advertising Restrictions on Cigarettes and Alcohol

We begin with the role of advertising on cigarettes. At issue is whether advertising can increase the total consumption of bads like cigarettes. The issues we will address are ones on which people differ and hold strong views. In a report on cigarettes, the surgeon general (1989) concluded at one point that:

> There is no scientifically rigorous study available to the public that provides a definitive answer to the basic question of whether advertising and promotion affect the level of

> tobacco consumption. Given the complexity of the issue, none is likely to be forthcoming in the foreseeable future.
>
> (pp. 516–517)

Even this conclusion was disputed by parties to both sides of the issue. Economists understand that well-intentioned interventions often have unintended consequences. Before blaming advertising for our ills, we should inquire into the nature of advertising.

THEORIES OF ADVERTISING Three main theories have developed about how advertising works and what it does for or to the community. Advertising alternatively is a form of information, a tool for persuasion, or a complementary good. The first two of these represent an old battle in advertising theory with contrasting villains and heroes: Information is generally beneficial, while persuasion is at least more questionable. The most recent addition treats advertising as a complementary good. Finally, Box 23.2 visits the advertisement of worthless goods—patent medicines.

ADVERTISING AS INFORMATION Nelson (1970, 1974) studied the implications of advertising as information. Viewed as information, advertising can be seen to lower equilibrium prices, create better access to the market for new entrants, and provide better matches of consumer preferences with feasible consumption bundles. Informed consumers find that their dependence on or loyalty to Brand A will be weakened by their improved knowledge of alternatives. If it is easy to opt for another brand, then the consumer has flexibility, and a flexible consumer is more likely to resist undesirable changes in the brand, such as a drop in quality or an increase in price. This greater responsiveness to price implies a more elastic demand, and it makes possible lower market equilibrium prices. How can the firm's costly advertising activity help but be passed on to consumers in the form of higher prices? The mistake is that while the price at a given output must rise, the market equilibrium quantity may change due to competition, and the equilibrium price may fall.

BOX 23.2

Can Advertising Lead Patients Astray? The Case of Medical Quackery

Excepting the medical heroes of history, such as Hippocrates and Galen, few doctors offered a scientific practice until the twentieth century. Modern medical science did not begin until the mid-nineteenth century and the typical patient confronting the typical physician did not attain even a 50–50 chance of getting better until about 1900. Most early practitioners may have had strong faith in their practices, such as bleeding the patient or applying strong purgatives, but these often may have worsened a patient's chances for recovery.

Oddly, there have also been many doctors throughout history who knowingly advertised and sold a worthless bill of goods. These "quacks" brazenly promised marvelous healing powers for such oddities as "patent medicines." Such practices persist today mixed with generally innocent "alternative medicines," herbal treatments. However, there still are modernized equivalents to the magical electric boxes and oddly designed "magnetizations."

The audacity of the advertising medical quack was not lost on Gilbert and Sullivan, the noted creators of comic opera during the late nineteenth century. Their views on quacks were captured by:

MIKADO:
My object all sublime
I shall achieve in time—
To let the punishment fit the crime,
The punishment fit the crime; . . .
The advertising quack who wearies
With tales of countless cures,
His teeth, I've enacted,
Shall all be extracted
By terrified amateurs.

ADVERTISING AS A BARRIER TO ENTRY In contrast to advertising as information, Bain (1956) argued that advertising differentiates one brand from another, creating increased brand loyalty. By making consumers more resistant to price changes and demand, advertising can result in greater market power and higher equilibrium prices. Adding to this vision, Comanor and Wilson (1974) showed that the persuasive power of advertising may cause it to be asymmetrically effective for the incumbent versus potential new entrants. Consumers have greater experience with established firms and greater recognition of them. The next advertising dollar to be spent by the established firm will yield a greater return than the same dollar spent by the newcomer.

ADVERTISING AS A COMPLEMENTARY GOOD Nobel laureate Gary Becker and colleague Kevin Murphy (1993) proposed a theory to account for these competing claims within a single model—one that appeals to an older theory of complements and substitutes. Let advertising be considered a good that is a complement to the good advertised. A consumer might wade through commercials with irritation during a ball game but enjoy the humorous one featuring a favorite beer. The commercial increases the consumer's marginal utility from consuming that brand and, under this theory, firm advertising will raise total consumption of the product, just as a reduction in the price of mustard will increase the consumption of hotdogs.

The Possible Effects of Brand Switching

When an imperfectly competitive firm advertises, it potentially improves its demand in part by inducing consumers to switch brands and in part by inducing consumers to consume more of the product in total. However, other firms also will advertise for these purposes, and the advertising among the firms may be partially or even totally offsetting. Advertising in the industry could have little or no effect on total industry demand, but it also could have a positive effect.

Laypeople often conclude mistakenly that the answer is obvious. Surely cigarette advertising must lead to more aggregate smoking or cigarette companies would not spend so much money on it. Cigarettes have been among the most heavily advertised products. A similar argument is made by some economists who note that the many available brands are really owned by a few companies and that the degree of brand switching (about 10 percent of

smokers switch brands in a typical year) is not sufficient to justify the billions spent on advertising (Tye, Warner, and Glantz, 1987). Sales figures confirm that 80 to 90 percent of U.S. cigarette sales are generated by the top four firms.

Others argue that brand switching is a sufficient motivation for heavy advertising. Two points help explain this view: (1) the firm's decision to advertise is made *ex ante* with available information, and (2) the firm's decision criterion is what would happen if it did not advertise. On the one hand, the firm commits money to advertising based on the effect it expects advertising to have. The firm may not have the econometric data published later, it may not believe what is published, or admittedly, it may have better information. On the other hand, the number of smokers who actually switch on average is less relevant than the unknown number who would switch if the firm did not advertise.

Increased Demand or Brand Switching?

To settle the question, we must examine the evidence. If we find that advertising increases total consumption, we could reduce consumption by restricting advertising. Studies suggest that cigarette advertising elasticities are very small. Research tends to find (Gallett, 2003; Baltagi and Levin, 1986) that advertising has no significant marginal effect on cigarette demand. Where the ad effect is found to be significant, the reported elasticities are generally quite small.[4] These elasticities fall between 0.1 and 0.2, and the most common result falls around 0.1. The Toxic Substance Board of New Zealand (1989), a strong advocate of advertising bans, conservatively uses an advertising elasticity of 0.07 in estimating the effect of an advertising ban.

Suppose that the true advertising elasticity were 0.1 exactly, and we applied this figure to estimate the effect of a total ban on cigarette advertising. We would reduce cigarette consumption by 10 percent, which is the product of a 100 percent price reduction and the 0.1 elasticity. Even this crude estimate might raise an unusual controversy in that ban advocates (arguing that the ban was successful) and ban opponents (arguing that it didn't change much) both might claim that this evidence is support for their case.

The elasticity estimates often are drawn from time trends of aggregate data, sometimes aggregated to the national level. At this level, little variation in advertising expenditures is captured, and the massive advertising levels outside of the cigarette industry, though relevant to cigarette demand, usually are ignored (Saffer and Chaloupka, 1998). It is also econometrically unsafe to extrapolate this far out of sample so we must study the effects of actual advertising bans.

World experience with advertising bans is informative. Several countries have banned cigarette advertising outright, and the United States installed a partial ban (on broadcast media) in 1971. The picture is clear regarding the U.S. experience. The 1971 ban of cigarette advertising on television and radio somewhat paradoxically may have increased cigarette consumption. Hamilton's (1972) account illustrates this peculiar history. Prior to the 1971 ban, an FTC ruling on the Fairness Doctrine required television and radio stations to give equal time to antismoking messages.

The result was a flurry of memorable commercials showing the harm of smoking. In one, a father appears walking in the forest with his young son. They stop to rest, lean back against a tree trunk, and the father reaches for a cigarette. His son watches in admiration as his dad lights up, but the father sees this, thinks, and stops. He gets the message and so do the viewers. Hamilton showed that these antismoking messages were effective in reducing smoking in the population, while the advertising expenditures had small, if significant, marginal effects. Thus, when the ban eliminated both procigarette and the mandatory anticigarette messages at the same time, the combined effect could have increased smoking.

Perhaps more important, the antismoking messages had a relatively greater effect on young, potential smokers. Lewit, Coate, and Grossman (1981) studied data on more than 6,000 youths. They found that television watching significantly increased the probability that a youth would start smoking, while the Fairness Doctrine ads reduced the probability that a youth would start smoking. Harris and Chan (1999) have also found the clear relation of price elasticity and age; from 15- to 17-year-olds through 27- to 29-year-olds, price elasticity in absolute value starts out high and continually declines.

For many years, the single cross-national study of the effect of advertising bans by Hamilton (1975) showed that the bans have no significant effect on cigarette consumption in the country. In a later study, incorporating many new years of data and experience with bans, Laugesen and Meads (1991) reported these bans to be effective. Studying data for 1960 to 1986 and developing an index measuring the advertising restrictions, their time series estimates showed that a country's banning of tobacco advertising will reduce tobacco consumption by 6.8 percent. Recent studies corroborate this result (Iwasaki, Tremblay, and Tremblay, 2006; Saffer and Chaloupka, 1998), though it warned that partial bans may invite tobacco firms to substitute advertising in unrestricted categories for banned advertising. Keeler and colleagues (2004) report that cigarette companies in the United States did exactly this in response to the U.S. tobacco settlement of 1998. Facing reduced revenues by 8.3 percent due to the consequent price increase, the companies offset about one-quarter to one-half of that through increased advertising.

Advertising and Alcohol Consumption

Although earlier studies reported no effect of advertising on alcohol consumption, Saffer (1991) found that such advertising was a significant factor in drinking. Similarly, Saffer and Dave (2003) found such advertising to be especially effective on youthful drinkers; they estimated that a complete ban on all alcohol advertising could reduce adolescent alcohol consumption by 24 percent, with even stronger effects on binge drinking. Taking a different approach, Saffer (1997) found alcohol advertising levels to be correlated positively and significantly with motor vehicle fatalities, *ceteris paribus*. He estimated that partial bans (bans on broadcast advertising) would reduce annual fatalities by 2,000 to 3,000, while a total ban could reduce fatalities by up to 10,000 nationally.

Excise Taxes and Consumption of Cigarettes and Alcohol

The public commonly believes that taxes on products are always and fully passed on to the consumer, but this is not true. The irony is that if it were true, then cigarette and alcohol taxes would have no effect other than to raise money for the government coffers. By first examining the theory of excise taxation, we learn the importance of the price elasticity of demand and supply in determining the incidence of the tax and the degree to which consumption is reduced. With this understanding, we will examine the empirical knowledge regarding these elasticities.

The Consumption-Reducing Effects of Excise Taxes in Theory

Excise taxes form another major tool that can reduce consumption of bads in populations. Figure 23.4 depicts the supply and demand for alcoholic beverages. If a sufficient number of

voters were to agree, the consumption of beer could be reduced from Q_E to Q_{E2} by imposing an excise tax increase of T dollars (recall that the tax shifts the supply curve vertically by the amount of the tax). The analysis also reveals that the tax would not fall entirely on the beer drinker, but the incidence would be shared. Assuming demand curve D_0, because the consumer pays a higher price after the tax, we say that the incidence on the consumer is $(P_{E2} - P_E)$. Because the producer pockets the lower amount, P_{S2}, we say that the incidence on the producer is $(P_E - P_{S2})$.

The research issue is illustrated by the alternative possibility that the demand curve for beer is D_1. The alternative equilibrium is at point A, at which the reduction in consumption (not shown) is less than the original reduction $(Q_E - Q_{E2})$. Thus, the reduction in quantity demanded will depend on which curve represents the true demand curve: the relatively inelastic curve D_1, or the more elastic curve D_0. Generally, the more responsive demand is to price (the greater the demand elasticity in absolute value), the greater the corresponding reduction effected by the excise tax increase.

From models of the consumer's choice over bads, such as the rational addiction model, we know to distinguish between long-run and short-run effects of price changes. We also learned to distinguish youth from adult consumers, a distinction with important policy consequences. Box 23.3 provides a further discussion related to tobacco policy.

Excise Taxes and Cigarette Consumption in Practice

The importance of price elasticity is not lost on economists, and most econometric studies of cigarette consumption report price elasticity estimates. Over the history of these studies, the

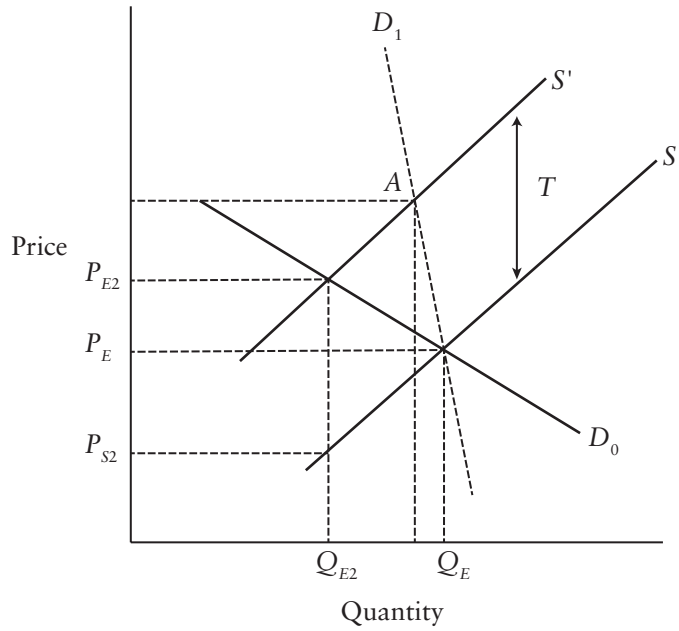


Figure 23.4 The Consumption of Alcohol Is Reduced by an Excise Tax Increase

reported cigarette price elasticities in absolute value range rather widely, from as low as 0.2 to occasional estimates greater than 1.0. There is no doubt that cigarette demand responds significantly to price, but the response is generally inelastic. Building on the earlier work, more recent studies find short-run price elasticities in a narrower band, typical of which is the 0.3 to 0.5 band reported by Keeler et al. (1993).[5] The United States Tobacco Settlement of 1998 effected a rise in the price of cigarettes, and consequently it offered a natural experiment on the effects of price on consumption. The results show a decline in cigarette consumption of 8.3 percent, though increased advertising by cigarette companies offsets this effect (Keeler et al., 2004). Sheu and colleagues (2004) estimated the price elasticity from these data to be in midrange, 0.46 in absolute value. More recent estimates by Carpenter and Cook (2007) support the previous findings of approximately 0.56.

BOX 23.3

Mind If I Smoke?

On a California billboard, the man's date responds to this question by asking: "Care if I die?" In a similar spirit, recent American films have depicted the inside of the tobacco industry as darkly menacing. The British news magazine, the *Economist*, ponders whether the tobacco industry has become "the new evil empire."

Meanwhile, other reports suggest that average smoking rates in the population continue to decline. The decline has also occurred among women, with an especially large drop in smoking among women of childbearing age. Tobacco companies generally have raised cigarette prices, and in California, the state legislature's cigarette sales tax increase at the beginning of 1999 was followed by an unexpectedly sharp decline in cigarette sales. The high level of legislative, activist, and consumer discussion and debate over tobacco continues throughout the country.

This scenario forms the backdrop for the tobacco companies' unprecedented November 1998 offer of $368 billion to the states in return for promises to limit further lawsuit activity. The American Cancer Society and the American Medical Association both gave their approval, albeit with some qualifications, and the settlement began with the highest hopes.

The legal theory applied in these contexts and to be used in future Department of Justice suits holds the tobacco companies liable for health costs incurred by the governments to provide care to cigarette-induced disease victims. The argument distinguishes cigarette issues from others involving voluntary risk-taking, such as skydiving, alpine skiing, or work in risky occupations. An important legal question is whether the tobacco companies deliberately withheld critical health information.

Early research found that youth were much more responsive to price increases and more accessible to the excise tax tool. This is an attractive result for anyone wishing to deter youth from taking up the smoking habit: "an excise tax increase . . . might continue to discourage smoking participation by successive generations of teenagers and young adults and gradually impact the smoking levels of older age groups as the smoking-discouraged cohorts move through the age spectrum" (Lewit and Coate, 1982, p. 143). Recent research tends to support the view that "price is a powerful determinant of smoking for high school seniors" (Gruber

and Zinman, 2002); they estimate the cigarette price elasticity of these youths to be 0.67. The findings of Sen and Wirjanto (2010) are much lower, however, at elasticities of 0.2 to 0.5, suggesting that this issue is still in dispute. Please be aware, however, that even low elasticities can do some significant good. For example, with 0.2 as the true number, a doubling of the price of a pack would reduce youth smoking by 20 percent. An effect on children you may not have expected: higher cigarette taxes tend to increase the BMI (body mass index) of children of smoking mothers (Mellor, 2011).

As predicted by some models, such as rational addiction, estimated long-run cigarette elasticities are larger than short-run elasticities in absolute value. Chaloupka (1991) applied the rational addiction framework and estimated long-run values approximately twice as large as for the short run. Keeler et al. (1993) found them to lie in a narrow range, 0.5 to 0.6 in absolute value. A more demanding test of the rational expectations model is one that takes seriously its claim that consumers make choices not only based on current cigarette prices but also on their ideas about future cigarette prices. The estimation process requires difficult econometrics, and ideal data sets are difficult to come by, but two recent studies confirm this forward-looking prediction (Baltagi and Griffin, 2001; Gruber and Koszegi, 2001).

With the elasticity magnitudes reported in these various studies, what possible effects could excise tax increases be expected to generate? To focus the discussion, suppose that a tax increase has been installed of sufficient magnitude to double the price of cigarettes. Assuming an elasticity of 0.4, the 100 percent price increase, by simple extrapolation, would reduce cigarette consumption by 40 percent, a substantial reduction. As we discussed earlier, where a similar extrapolation for advertising bans suggested a more modest reduction, such out-of-sample extrapolations are risky, but the example serves to illustrate the potency of the tax tool. Furthermore, the elasticity estimates for youth are large enough to double this effect for that age cohort. Likewise, the larger long-run elasticities suggest the greater policy effects as time passes. Direct measures of the effect of cigarette taxes on mortality also show the effectiveness of tax policy. Moore (1995) tested tax variables in equations to predict mortality from several smoking-related diseases. Higher taxes significantly reduce mortality from lung cancer, cardiovascular disease, and asthma. He reports that a 10 percent cigarette tax hike would save 3,700 lives per year in the United States. Evans and Ringel (1999) studied the effect of cigarette excise taxes on birth outcomes. Their research found the taxes effective in reducing smoking by expectant mothers and generating better, higher birth weights in the newborns.

Although some countries have applied cigarette taxes vigorously to reduce population smoking, the United States has not done so until recently. In 1997, the U.S. average tax rate, at 35 percent of the average price of a pack, was the lowest among 29 countries in one study. The United Kingdom, Ireland, Germany, and all of Scandinavia ranged from 70 percent to 85 percent. The U.S. federal excise tax declined in real terms after the surgeon general's mid-1960s announcement that cigarette smoking causes cancer until recent increases to $0.39 per pack (in percentage terms it was still lower in 2002 than it had been in 1969). Figure 23.5 reveals that the tax was reduced by allowing a nearly constant nominal tax rate to be eroded in percentage terms by the secular rise in prices per pack.[6]

MEDICAL AND SOCIAL INFLUENCES ON SMOKING While health economists often emphasized prices, taxes, and advertising as influences on demand for health bads, there has been a growing interest in medical and social factors. Saffer and Dave (2005) find that people with a history of mental illness are much more likely than the average person to consume alcohol (26 percent more likely), cocaine (66 percent more likely), and cigarettes (89 percent more likely).

Social influences also play a strong role. DiCicca and colleagues (2006) found that an index of local attitudes toward smoking was related to cigarette consumption levels. The

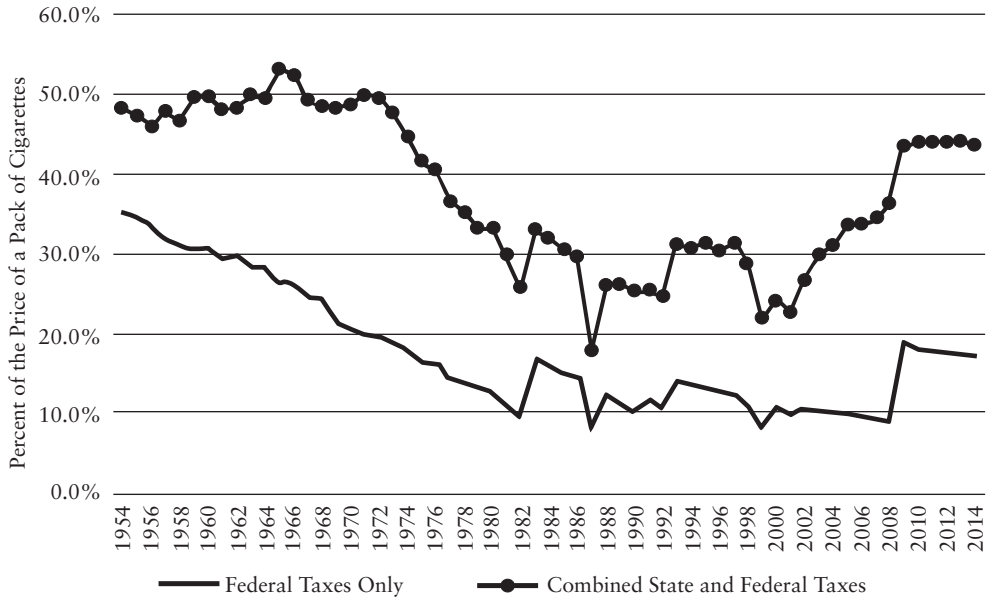


Figure 23.5 State and Federal Taxes as a Percent of the Price per Pack
Source: Orzechowski and Walker (2014).

more adverse were the antismoking sentiments toward smoking in the local culture the lower were the smoking levels. Studies by Folland (2008) and by Brown et al. (2006) find that individuals with a strong social network are less likely to smoke. Other studies focus on teenagers, finding that these adolescents are strongly influenced by their social peers regarding smoking behaviors (Katzman, Markowitz, and McGear, 2007; Clark and Loheac, 2007).

Excise Taxes and Alcohol Consumption

Studies of alcohol consumption, price, and advertising have often focused on youth; these younger age groups exhibit the highest rates of alcohol abuse, such as binge drinking (Cook and Moore, 2000). The beer tax in Figure 23.4 was depicted as substantially effective in reducing consumption, but would this effect be experienced in real life? Grossman et al. (1998) find that young adults respond to beer price increases, and the elasticities range from 0.2 to 0.4 (in absolute value) in the short run, with long-run elasticities 60 percent higher; other beer elasticity estimates have ranged even higher. Studies of wine and spirits yield elasticities ranging from 0.3 to 1.8 (Saffer and Chaloupka, 1998). More recent results by Sen and Campbell (2010) provide more somber evidence. They find that license-related regulations and excise taxes on alcohol significantly reduce motor vehicle fatalities involving children.

The latter study also reports substantial, though inelastic, estimates for price elasticities of illicit drugs including marijuana, cocaine, and heroin. Thus, taxes should have major impacts on consumption of addictive substances.

ILLEGAL DRUGS AND PROHIBITION While it is doubtful one would find people who wish to prohibit smoking entirely or to return to the 1920s prohibition against alcohol, the subject of prohibition is highly relevant to consumption of illegal drugs. America's "War on Drugs"

serves as an example. Should this War on Drugs be modified or stopped? We leave the politically contentious policy issues to a more policy oriented venue and focus instead on a theoretical issue raised by Becker, Murphy, and Grossman (2004). Which is better, criminalization with enforcement or legalization with an optimal excise tax?

The authors first make the point that when the consumption demand for the drug is price inelastic, increases in enforcement can cause substantial increases in the money that drug smugglers spend to distribute their product to consumers. These extra expenditures are largely counter to social welfare; they must avoid the police, fight off competitors, and distribute the drugs by stealth in criminal networks. Figure 23.6 illustrates this point. Becker and colleagues make the simplifying assumptions that the drug dealers are perfectly competitive and produce at constant unit costs. We associate the unit cost levels, $C(0)$ and $C(E)$, with the legalization case and the criminalization case respectively. When the drugs are legal, there are no police enforcement expenditures and correspondingly no extra expenditures by the drug firms to avoid prosecution, so $C(0)$ indicates unit costs when enforcement costs are zero. In contrast, when the drugs are illegal, the unit costs will be higher as drug firms must pay not only production costs but extra costs to avoid prosecution and to maintain illegal distribution networks; in this case, the unit costs are $C(E)$, where E equals the enforcement costs. For these reasons, $C(E) > C(0)$.

Suppose that drugs are legalized so that $E = 0$, and the market equilibrium is at F. Drug seller costs are for production and distribution only, represented by the area of $0DFB$. In other words, all costs go to production and distribution. Contrast the case where enforcement costs, E, are positive and the smugglers must evade the law. The new market equilibrium is at G. Note the implication is that smuggler costs have become the area $0CGA$.

These become substantial and the reduction in consumption, $B - A$, is quite small. The observation by eye that $0CGA > 0DFB$ is corroborated by microeconomic principles which state that under inelastic demand a rise in price will increase revenue. Why is the familiar result for revenue relevant here? This is because the authors' assumption that

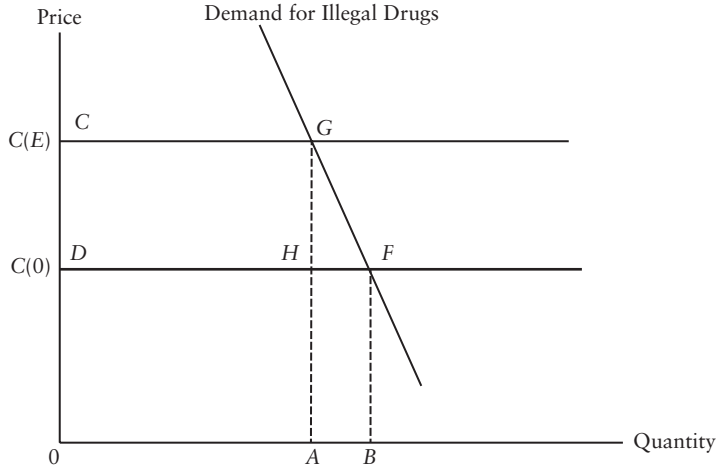


Figure 23.6 Comparing Prohibition versus Legalization under Conditions of Inelastic Demand

the smugglers are competitive implies that in the long run the firms' costs will equal their revenues. In other words, since we know that revenues have increased and since the firms' revenues equal their costs, we know that drug firm costs have increased, and realistically the illegal drug firm costs will be substantial. Recall that the illegal drug firms' costs are in large part spent avoiding prosecution and related criminal activities. Costs like this are unlikely to provide a social value, and most people would say that they are very harmful to society.

Contrast how society seeks to reduce drug use under legalization. The most prominent economic tool of government would be to institute an excise tax on the drug. While not shown here, tax revenues always accrue to the government, and these can be used to improve social welfare. The optimal excise tax is familiar to students of microeconomics: one installs a tax high enough to equal the marginal external cost to society of the drug consumption. Although the reader must refer to Becker, Murphy, and Grossman (2004) for the analytic development, it is well known from the economic theory of externalities and market failures that a properly chosen excise tax can in principle improve social welfare.

In Figure 23.6, an excise tax would be represented by a horizontal line. Becker and colleagues, through an analytical welfare analysis, addressed the question: Would the optimal price (with tax) under legalization be higher or lower than the unit costs under criminalization with optimal enforcement costs, *E*? Given the context described in this section, the authors concluded that the price (with optimal tax) under legalization would be higher than the price of illegal drugs under criminalization. This result would make a strong case for legalization. With higher drug prices consumption would decline: a benefit to society. Second, it would reduce or eliminate the criminal activities and expenses created by illegal drug suppliers.

Conclusions

This chapter examines the economic nature of health bads, and it studies the potential of curbs on advertising and increased excise taxes to reduce the consumption of bads. Econometric estimates of the effect of advertising for cigarettes tend to report small and sometimes insignificant elasticities. These are estimates of changes at the margin. In contrast, advertising bans entail by definition large reductions in advertising levels; the most recent work reports that bans have some significant effect when studied on an international basis. Related studies on alcohol advertising also report statistically significant effects.

Excise taxes, however, appear to be more potent. Though cigarette and alcohol price elasticities are in the inelastic range, they are large enough in absolute value to have substantial potential as a curb to consumption if the public chooses to apply them. Consistent with rational addiction models, the long-run price elasticities are greater in absolute value than the short-run elasticities. Cigarette price elasticities for youth tend to be larger in absolute value than those for adults.

Summary

1 The nature of addiction, though as yet unresolved, generates behaviors that appear generally consistent with a rational addiction model. Irrational or myopic addiction models provide viable alternatives.

2 Researchers generally assume addictive behavior to involve both "reinforcement" and "tolerance." A myopic, or near-sighted, addict looks solely at the reinforcement effect. A rational addict, however, considers the future harmful consequences of current addictive behavior.
3 There are three contrasting theories of advertising. In one, advertising is primarily information, which leads to greater competition and possibly lower prices. The second describes advertising as a potential barrier to entry, which increases monopoly power and prices. The third describes advertising as a complement to the good advertised; it increases marginal utility of the product.
4 Advertising bans and related restrictions appear to have significant effects on consumption of cigarettes and alcohol. Though total effects of advertising are small in econometric studies, they are somewhat more substantial in a recent international study of total bans.
5 Excise taxes work to curb consumption in theory by increasing the price of the product. Then, the effect on consumption depends on consumer responsiveness to price, which is the price elasticity.
6 Increased excise taxes on cigarettes and alcohol appear to be more potent public policy tools for curbing the consumption of bads than are advertising bans. Though estimated price elasticities are small in absolute value, they can be combined with substantial price changes to result in correspondingly large reductions in consumption.

Discussion Questions

1 Many students smoke, drink alcohol, or ingest other addictive substances, such as caffeine (in coffee, tea, or soft drinks). How do the addiction models that are presented relate to students' everyday habits?
2 Do cigarette and alcohol ads you have seen contain primarily informative content or persuasive content? What consequences would you predict from your finding?
3 Someone says: "The advertising elasticity of cigarette demand may be small, but it is big enough to warrant policy to ban cigarette advertising." What would "big enough" mean in this context?
4 What does the evidence on the effects of the partial U.S. ban on cigarette advertising suggest about the relative effectiveness of cigarette advertising versus antismoking advertising?
5 How does the responsiveness to cigarette advertising and price differ between youths and adults? Why is this important?
6 Explain how it is possible to advertise too little or too much.
7 Suppose it takes considerable time for the large majority of cigarette smokers to become fully informed about the fact that cigarette prices have risen all over. How would this matter to the effectiveness of a tax hike? What other effect of a cigarette tax hike may take considerable time?
8 How would an excise tax hike in Kansas be helped or harmed in its attempt to curb smoking among Kansans if the neighboring states (e.g., Missouri, Nebraska) did not also hike their cigarette excise taxes?
9 When all costs are considered, which public policy tool for curbing the consumption of bads would be the most costly to administer for a given amount of consumption reduction among consumers?

Exercises

1 The analysis accompanying Figure 23.3 investigates the impacts of an increase in cigarette prices. Use the figure to show both the short-run and the long-run impacts of a cigarette price decrease.
2 If the elasticity of aggregate cigarette demand with respect to advertising were 0.15 in absolute value, by extrapolation what effect on cigarette consumption would be caused by a 10 percent reduction in advertising? A 50 percent reduction? A 100 percent reduction? How and why does one's confidence in prediction change over this range of reductions?
3 Suppose the price elasticity of cigarette demand is 0.4. If we increased the prices of cigarettes by 50 percent, what would we expect to happen to the quantity purchased? To total expenditures on cigarettes?
4 In the discussion on rationales for intervention in markets, we note that Manning found external costs of $0.33 per pack of cigarettes.
   (a) Draw a supply and demand diagram, and graph Manning's external costs of $0.33 (in 1995 dollars) based on a market price of $1.50 (in 1995 dollars) per pack.
   (b) If a tax of $0.33 were imposed, what would happen to the market price, and to the equilibrium quantity?
5 Using Manning's estimate of external costs of $0.33 per pack of cigarettes:
   (a) Calculate the new market price for cigarettes using a demand elasticity of 0.25. Would this be economically efficient?
   (b) Calculate the government's revenue from a $0.33 tax on cigarettes in part (a).

Notes

1 In addition, cigarettes are implicated in low birth weights (Rosenzweig and Wolpin, 1995).
2 We can show that the *A* curve must cross the *C* line from above for an equilibrium to be stable. This occurs at point *D*. It does not occur at point *E*.
3 See Chaloupka and Saffer (1992) and McCarthy (1993).
4 McGuiness and Cowling (1975), Seldon and Doroodian (1989), and Tremblay and Tremblay (1995). Roberts and Samuelson (1988) simultaneously estimate brand switching and total consumption effects, concluding for their data that total consumption effects dominated.
5 See also Tremblay and Tremblay (1995). Smokeless tobacco demand also responds to price in econometric studies, both on price and the price of substitutes (Ohsfeldt, Boyle, and Capilouto, 1998).
6 In 2009, the federal excise tax jumped from $0.39 per pack to $1.01 per pack, an increase of $0.62 or 159 percent. It has not changed since.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 24

The Economics of Social Capital and Health

In this chapter

- What Is Social Capital?
- How Do People Choose Social Capital?
- Empirical Tests of Social Capital and Health
- Pursuing Causality
- Elements of Trust
- Social Capital and Risky Choices
- Conclusion

What Is Social Capital?

The benefits from positive social relationships, both to individuals and communities, were understood by the Greeks, as well as centuries later by Enlightenment authors. It is ironic that a good idea which survived so long through history should only recently come to be a study of science.

Researchers from several academic disciplines have developed the ideas more fully, including economists (Loury, 1977), sociologists (Coleman, 1988; Bourdieu, 1985), political scientists (Putnam, 2000) and epidemiologists (Kawachi, 1999). Though each discipline defines social capital somewhat differently, there are strong similarities with Putnam's single statement: "Social capital refers to connections among individuals in social networks and the norms of reciprocity and trustworthiness that arise from them" (Putnam, 2000, p.19).

This chapter describes and applies the theory of capital with social capital. The second part introduces the Social Capital and Health Hypothesis, explains the empirical exploration, and follows the steps of progress of econometrics to confirm the correlation studies and then by applying certain "crucial tests" to conclude that social capital is causal to improved health.

How Do People Choose Social Capital?

The Individual Case

Social capital investment is chosen over many periods within a complex environment. As an introduction to a full model, we start with a more familiar and simple single period maximization problem. In this single period, the individual social capital, S, that John can enjoy is the amount that he gathers through socializing and participating in the community, which we call his investment, I. For this special case, we will assume that S equals investment I. John's task is to maximize profit in equation (24.1).

$$\Pi = S \cdot R(\hat{S}) - w \cdot C(I) \tag{24.1}$$

Here John chooses an amount of social capital, S, that maximizes his net income or profits, Π, where R is his perceived reward in dollar terms for each unit of S. We assume that R increases when the community average social capital, $\hat{S}$, becomes higher. $C(I)$ measures John's cost in terms of hours required to gather the S. Finally, w is his opportunity cost per hour spent on social capital.

Maximizing John's profit requires that marginal revenue, $R(\hat{S})$, equals marginal cost, $wC'(I)$.

$$R(\hat{S}) = w \cdot C'(S) \tag{24.2}$$

The full model expands on this, but at present this introduction makes it easy to find John's response to parameter changes.

Let R increase so that $R(\hat{S})$ shifts upward from $R(\hat{S})_0$ to $R(\hat{S})_1$ and this implies that John invests more from I_0 to I_1. The logic of an increase in $\hat{S}$ is similar. Alternatively, let marginal cost, $C'(I)$, fall or let w decrease. As shown, this increases investment from I_0 to I_2. To summarize John's investment response to parameter change:

1 Increased R ➔ increased investment
2 Increased $\hat{S}$ ➔ increased investment
3 Decreased C' ➔ increased investment
4 Decreased w ➔ increased investment

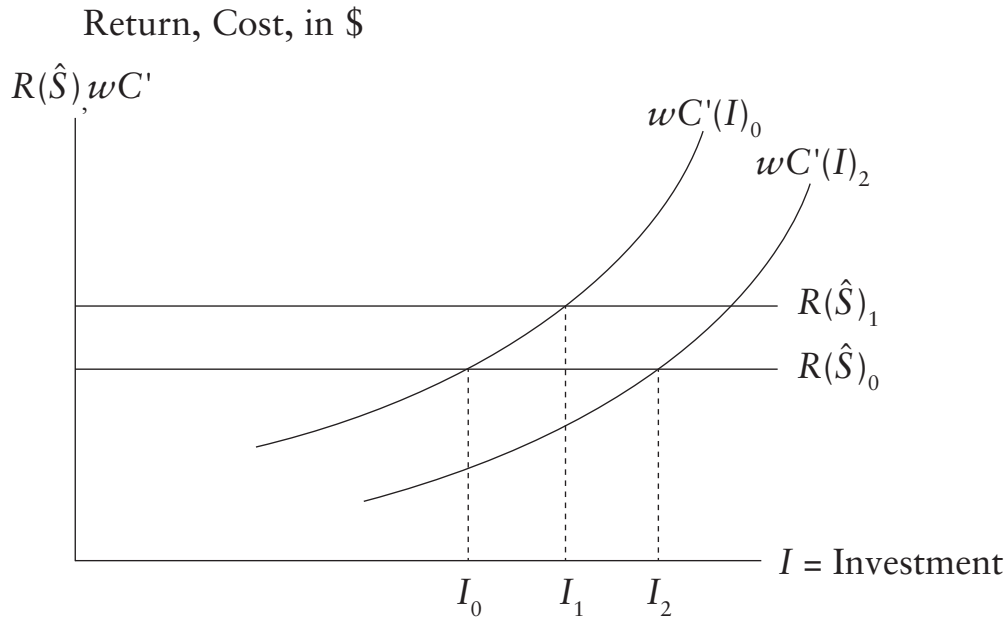


Figure 24.1 The Effects on Investment of Change in the Parameters: R, $\hat{S}$, w, and C

Glaeser, Laibson, and Sacerdote (GLS) (2002) developed a full multi-period social capital investment model where investments are chosen for each period. Equation 24.3 is the full model and is presented for demonstration purposes.

$$\max_{I_0, I_1, \ldots I_T} \sum_{t=0}^{T} \beta^t [S_t R(\hat{S}_t) - wC'(I_t)] \quad (24.3)$$
$$\text{Subject to the constraint } S_{t+1} = (1 - d) S_t + I_t$$

More features are included in the full model: The subject line depicts how social capital progresses through time with investment and depreciation. There are three new parameters: t is John's age; β is a fraction that describes how seriously he considers future periods; and d is the depreciation rate for social capital. The heart of GLS is the profit function, in the brackets, which takes the same form as equation (24.1). The effects of parameter changes are also found in the same way. John asks, "How does a given parameter change affect the profitability of my investment in social capital?"

We will not work through the mathematics of the GLS model, but the results confirm our introductory model in equation (24.1). The following parameter change effects are seen as a response to the GLS model:

1 Increased R ➔ expanded investment
2 Increased $\hat{S}$ ➔ expanded investment
3 Decreased C' ➔ expanded investment
4 Decreased w ➔ expanded investment

These four are exactly the same as the introductory model. Students do not need to solve the GLS model and are not asked to do so. However, these are the remaining reported solutions:

5 Increased t ➔ lower investment
6 Increased β ➔ expanded investment
7 Increased d ➔ lower investment

Theory like GLS helps to explain what we see. For example, John's perceptions of his rewards, R, determine his investments in S. If science findings show that social capital benefits his health, this would likely raise his R and stimulate his social capital investment. The theory also generates other empirical testable hypotheses. For example, as John ages and t rises, his investment in S declines. This hypothesis was already tested and supported by GLS. Finally, GLS describes an intriguing feedback mechanism between the community average social capital, $\hat{S}$, and John's perception of the rewards. We pursue this and other community social capital ideas next.

Community Social Capital

In early decades of social capital research, the focus was on the community. Robert Putnam's (1993, 2000) empiricism found unmistakable signals of social capital benefits from the health data. He found that U.S. states with higher $\hat{S}_t$ tended to have healthier children, lower murder rates, better education performance, more tolerance of other ethnicities, and better health status (which becomes the main focus of this chapter's empirical reports). What can a community do to improve its $\hat{S}$?

The GLS model suggests the value of lowering the cost, $C(I)$, of a given level of investment for all residents. Mayors and other administrators seek to provide amenities to city dwellers, if only to improve the odds of re-election. The most beneficial to social capital are those which better enable people to meet and socialize: city parks, walkable sidewalks, efficient transportation, adequate street lighting, effective policing, participatory governing, and volunteering opportunities. Urban planners know that these are valued by residents as amenities, but the task for health economists is to find whether or not these gains in social capital also generate gains in people's health.

How Could Increments to Social Capital Improve Health?

There are at least four pathways for social capital to result in health gains: (1) Stress reduction: many diseases are stress related: asthma, ulcers, psychological problems, depression, sleeping disorders, and possibly many others. An unusual look at stress is offered by the scientist Sapolsky (1998), who found that humans dwell on their stress much longer than do animals like zebras: Their stressful events are life or death shocks; the lion either gets you or he doesn't. However, when it is over the zebra quickly returns to normal. (2) Information: social groups would know and encourage healthful practices such as moderate drinking, no smoking, and knowing where to get appropriate medical care. (3) Responsibilities: social ties are valued directly for themselves as they provide utility. This alters the individual's desired rate of tradeoff between risks to life and health and the rewards. (4) Social groups will be more effective than individuals at encouraging development of healthcare in a community (Anderson, 2004).

Students should understand that there is no mathematical necessity that every social group will provide positive influences on its members. Teenagers may be introduced to cigarette smoking by their peers. A study of small German towns in the 1930s found that social groups promoted fascism (Satyanath, 2013). Even families may be dysfunctional, but if such contrary

cases predominated they would show up in the empirical support or lack of support for the Social Capital and Health Hypothesis. We must let the data decide.

Empirical Tests of Social Capital and Health

Researchers measure social capital S in two ways, by community characteristics or by individual characteristics. Harvard political scientist, Robert Putnam, earned the credit for jumpstarting empirical research on the effects of social capital. Putnam developed a measure of community social capital, $\hat{S}$, as an aggregation of 14 variables, with each indicating the extent of participation and trust in the community. These are listed in the note below.[1]

The scattergrams in Figures 24.2 and 24.3 reveal a strong beneficial association between $\hat{S}$ and the state health index and with the state's mortality rates.

Economists contribute to this interdisciplinary field by introducing economic modeling and multivariate econometrics.

Figures 24.2 and 24.3 strongly encourage further exploration, but these correlations show only two variables at a time. We need to test whether the $\hat{S}$ result survives when other plausible variables are included. This test is emphasized in economics for strong reasons. To illustrate, at one time everyone thought that better nutrition was a major source of population growth from the 1700s on (Mckeown, 1976). Studies of these two variables gave strong support, but there were other important variables being omitted: improvements in public health, cleaner water, scientific medicine, new technology, and more. Robert Fogel (2011) studied the many other possible variables together and concluded:

> In many parts of the world, including the United States in the 20th century, medical advances appear to be at least as important as improvements in nutritional intake.

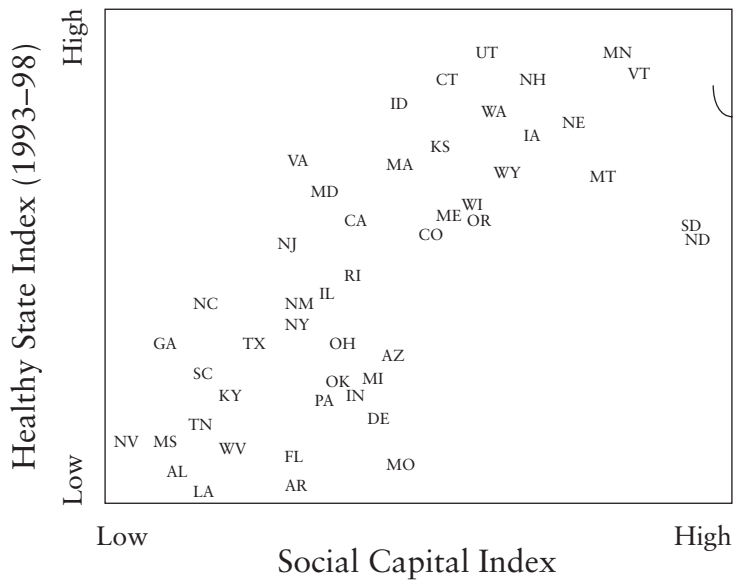


Figure 24.2 Population Health Is Better in High-Social-Capital States
Source: Putnam (2000).

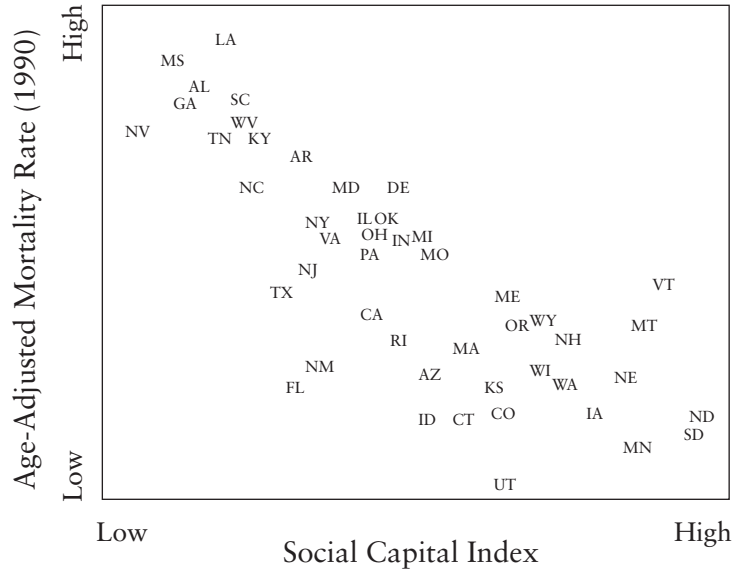


Figure 24.3 Mortality Is Lower in High-Social-Capital States
Source: Putnam (2000).

Table 24.1 (Model 1) presents regression results that measure the contributions of many variables—in predicting mortality rates while using Putnam's social capital definition. For students unfamiliar with regression analysis, we offer a refresher in Chapter 3. The coefficient in the following Table 24.1, when negative, indicates that the variable lowers the mortality rate. For example, here the *community social capital*, $\hat{S}$, coefficient is –480.7 in the mortality rate equation and we infer that social capital lowers the mortality rate.

Interpretation for other independent variables may require more effort. For example, it may be puzzling that higher income can make one sicker, as suggested in the table. Pritchett and Summers (1996) found that "richer is better," but this applied to third-world countries and we know that poverty hurts health in developed countries too. Deaton and Paxson (2001) find that high incomes could harm health in the United States, and perhaps it depends on what wealthier people spend their wealth on. Note that college education, which we know benefits income, also benefits health. Assessing the progress thus far, Putnam's measure of community social capital, $\hat{S}$, passes the test very well. The Social Capital and Health Hypothesis is supported so far. Other published studies find similar community social capital and health results: Folland (2006); Brown et al. (2006); Folland (2007); Scheffler et al. (2007); Brown et al. (2011).

Having seen that community-level studies tend to support the social capital hypothesis we ask whether individual studies would support it as well. D'Hombres et al. (2010) provide an example of how social capital, *S*, performs at the individual level.

For Table 24.2, Model 2, the dependent variable is "self-reported health," and the authors define the individual social capital in three parts: (1) *Trust*; (2) *Membership*; (3) *Social Isolation*. *Trust* is set to 1 (else 0) if the subject "greatly or quite strongly trusts a majority of people." *Membership* is set to 1 (else 0) if the subject is a member of a local organization

Table 24.1 (Model 1) The Effect of *Community Social Capital, CSC*

Independent Variables	*Total Mortality Rate*	*Infant Mortality Rate*
Community Social Capital (CSC)	–480.7***	–5.653**
% with BA Degree	–8.23**	–0.166**
Per Capita Income	0.011**	0.0004**
Poverty Rate	491.9**	17.95**
Unemployment Rate	0.0185	–0.001
Healthcare Expense per Capita	–0.345**	–0.016**
R Square (p for F)	0.609 (0.000)	0.715 (0.000)

Notes: * significant at 10%; ** significant at the 5% level; *** significant at the 1% level.

These regressions were based on a six-period panel measured in four-year intervals from 1978–1998 covering the contiguous United States. *Community Social Capital*, *CSC*, is an aggregate of indicators available in the Chicago DBD marketing data; these are subject responses to "which of the following apply to you": a. attended club meetings; b. helped in community projects; c. entertained at home; d. volunteered; e. "people are generally honest"; f. visited friends.

Table 24.2 (Model 2) Self-Reported Health and Individual Social Capital

Independent Variable	*Least Squares Regression*	*Instrumental Variables Version*
Trust	0.068**	0.079**
Membership	0.002	0.251*
Social Isolation	–0.115**	–0.228**
Age	–0.008**	
Sex dummy (female)	–0.099**	
Primary education	–0.090**	
Tertiary education	0.044**	
Work status (employment)	0.029**	
Minority status	–0.007	
R squared	0.15	
Observations	11,187	

Notes: * significant at 5%; ** significant at the 1% level.

Other independent variables were included in the regressions but are not shown in the table: household conditions (economic, material, size, number of working members); healthcare facilities (distance from doctor, distance from hospital, water quality, population size, road); other (village dummy, capital dummy, county dummies). Note also, the authors report the instrumental variables on the table. The instruments included *indication of community heterogeneity* and *description of county social classes*.

for church, art, sport, music, and so on. *Social Isolation* is set to 1 (else 0) if subject "feels alone." The sample is unique, it combines data from Armenia, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Russia, and Ukraine, all countries in transition from the Soviet period.

Table 24.2 (Model 2) shows that the Social Capital and Health Hypothesis is supported by the results for *Trust* and *Social Isolation*, but *Membership* has no significant effect. Why not *Membership*? Recall that the GLS model found that *Membership* tends to taper off as the subject ages. *Social Isolation*, a severe reverse measure of social capital, has the expected negative result. *Trust*, ubiquitous in social capital research, generally benefits the subject's health, and performs well here. The mysteries of *Trust* will be discussed shortly.

Model 2 shows that social capital "works" at the individual level and it works when *S* is identified in ways more common to health economics. Finally, since these eight post-Soviet (2001) countries differed substantially in political and cultural history from the developed West (e.g., OECD), the results suggest that social capital has an international effectiveness.

Pursuing Causality

Are the results from Model 1 and 2 sufficient to establish the "Social Capital and Health Hypothesis"? Unfortunately no. Health economists still need to identify whether social capital *causes* the increased health. Correlation is not the same as causality.

In science, one develops a treatment group, which is given the experimental effect, and one contrasts their response with a control group, which has not. In health economics, where the focus is on free human economic behavior, we often attempt this through econometrics. On rare occasions, useful treatment and control groups may appear in nature, sometimes from historical accidents—these are called natural experiments. These are rare but that does not rule them out; they are worth understanding.

Natural Experiments

The most famous natural experiment comes from epidemiology and provides the best illustration. In 1854, London was hit by a severe epidemic of cholera. The physician, John Snow, theorized that cholera was borne by water. He pursued the source, as a modern-day epidemiologist would, by mapping the Soho area, marking each spot where a cholera incident occurred. A pattern emerged that showed a concentration around Broad Street, where there was a pump that people used for water. Legend has it that John Snow broke the handle of the Broad Street pump and stopped the epidemic. Given John Snow's strong science reputation and this dramatic use of it, many today consider him to be the father of epidemiology.

Testing the Social Capital Effect for Causality

The following are two prominent econometric approaches.

OMITTED VARIABLES We can mistake our regression results for *S* if some other variables interact with both *S* and *H*, so as to fool us. For example, suppose that college students' average health status, *H*, exceeds that of the average American (it does) and that they form social contacts, *S*, more often. This could cause *S* and *H* to be positively correlated statistically, and yet it would be a mistake to conclude from this that *S* causes *H*.

One solution is to add education and other variables to the regression, as was done in Model 1 "*Percent with BA*" and in Model 2 as "*Tertiary education*." A related issue is to find if yet more variables would matter. Researchers try adding all seemingly relevant variables; one does not know which *a priori*. In Models 1 and 2, we have shown the social capital hypothesis to look stronger but causality still needs a more crucial test.

REVERSE CAUSALITY AND THE CRUCIAL TEST The key hypothesis that *S* causes *H* has been supported thus far, but suppose instead that causality runs in reverse, that *H* causes *S* instead. The correlation alone cannot tell us which is which. This possibility is not farfetched. Those who are healthy to begin with are more likely to be mobile. They can walk around, visit friends in the neighborhood, drive to shops and cafes, meet people, and perhaps participate in sports. They can develop social bonds with the people that they meet. In this case, good health has caused social capital to increase. This is reverse causality.

The crucial test for this is called "Instrumental Variables" or "IV." Provided that available data meet the required tests, this method can determine if the observed increase in *H*, the health variable, was in fact caused by the increase in the social capital variable. Models 1 and 2 have both been studied by this method. Applying the method to Model 1 found that $\hat{S}$ continued to show a significant negative effect on total mortality, though it lost its negative effect when the method was applied to infant mortality. Model 2 reports the authors' IV results which are seen on Table 24.2. These show significant *S* effect for each social capital variable; each also takes the right sign. Since these two studies, other causality studies have been done including Brown et al. (2010) and Ronconi et al. (2010). These too are generally supportive of the Social Capital and Health Hypothesis.

Might there be some reverse casualty happening at the same time? Both Sirven and Belmond (2012) and Rocco et al. (2011) found evidence of this simultaneity and report that the causal relationships between social capital and health go in both directions.

These successful causality results are key supports of the science behind the growing number of academic disciplines advocating improvement in social capital. For health economists and probably many others the support for the Social Capital and Health Hypothesis is a very important step.

Elements of Trust

Trust plays a key role in both social exchanges and economic exchanges, as recognized by Kenneth Arrow (1972), whose works formed the wellspring for health economics. In economic exchange it means that your word can be counted on for price and quality. Even the Vikings, during their most violent era, were successful traders. As used in the economics of social capital, the meaning of trust extends to a wider sense of acceptance.

The *Trust* variable as typically applied in these studies is usually defined by the subject's response to the question: "can people generally be trusted?" on a scale from 1 to 5, 5 being a strong yes. It is interpreted as an indicator of the underlying degree of a person's level of positive experience with the people he or she meets. We trust people we find "trustworthy."

Note that variables like *Trust* and *Self-Reported Health* are perceptions of the subject. Some studies prefer more objective measures like mortality rates, family status, size of social network, the subject's community participation, or memberships. The results of these are often similar to those reported here, though sometimes more mixed.

The Geography of Trust and of Social Capital

The World Values Survey reports average trust values for each country in Europe, revealing a north to south pattern. The latitude of each country's capital is highly correlated with trust (0.80), that is, there are higher trust values the further north.

The same north–south pattern appears in the United States. Before concluding that "colder is better" for developing trust, a better hypothesis may be that ethnic homogeneity increases trust. Figure 24.4 shows a strong positive correlation between ethnic homogeneity (measured by the Herfindal Index) and the average level of trust. People are more likely to trust other people who "look like them."

This poses a daunting challenge for policy in this world era of high migration, since it is easier to undo population homogeneity than to create it. A study by Putnam (2007) found that when immigration lowers the homogeneity of an area, the social capital measures are reduced. Yet, American experience illustrates the rewards of immigration and suggests that the reduction in social capital may be a temporary effect. The Irish, Jews, African-Americans, Roman Catholics, Italians, and many other groups were greeted on coming to America with both welcome and some hostility. However, they improved American culture, education, medicine, science, economics, and the professions.

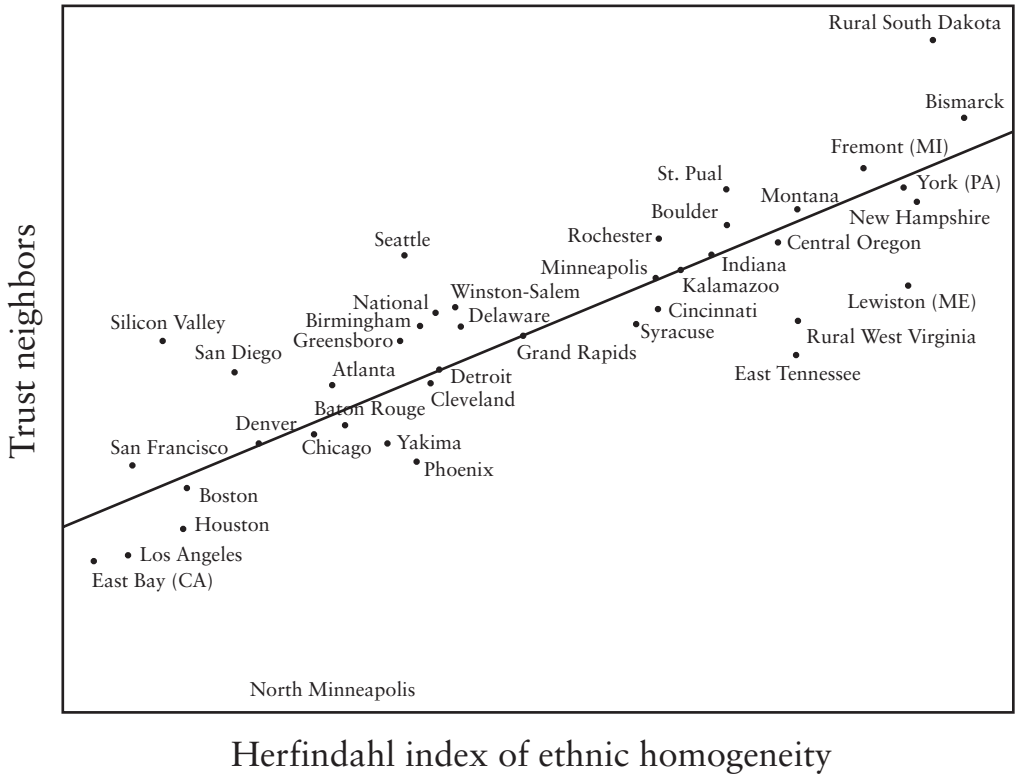


Figure 24.4 Illustration of the Positive Correlation between Ethnic Homogeneity and Trust in Neighbors

Source: Putnam (2007, p. 148).

Social Capital and Risky Choices

The social capital model of risky behavior is fairly clear: the more S, the more one can lose. First, accept two simplifying assumptions: (1) let S be determined exogenously; (2) let the other asset, money, be valued independently of S. Jack's utility is $U(S, m)$, while the risky gamble is made to acquire a quantity of cigarettes accompanied with a risk of death at rate p. Jack may alternatively consider p to be the risk of ill health. His expected utility then is

$$\textit{Expected utility} = (1 - p)U(S, m) \tag{24.4}$$

Jack maximizes expected utility subject to a market offer constraint. As shown in Figure 24.5, this is a concave curve $m(p)$. These are not merely interesting facts to him; these are the constraints, because cigarette smoke unavoidably creates a specific risk. The height of the curve $m(p)$ is the money metric of the pleasure he gets from smoking the cigarettes. As shown, indifference curves over (p, m) are convex and slope upward, given that risk is a "bad." His initial equilibrium is shown at E.

Suppose then that Jack and Jenna marry, giving a discreet increase in Jack's social capital S. The expected value of his life, as measured in equation (24.4), has increased. Consider the effect on his indifference curves.

They are steeper. The slope, or steepness, of an indifference curve measures the rate at which the subject is willing to trade one good for the other and still remain indifferent. In Figure 24.6 Jack's indifference curve I_1' is steeper than I_1 and goes through the original

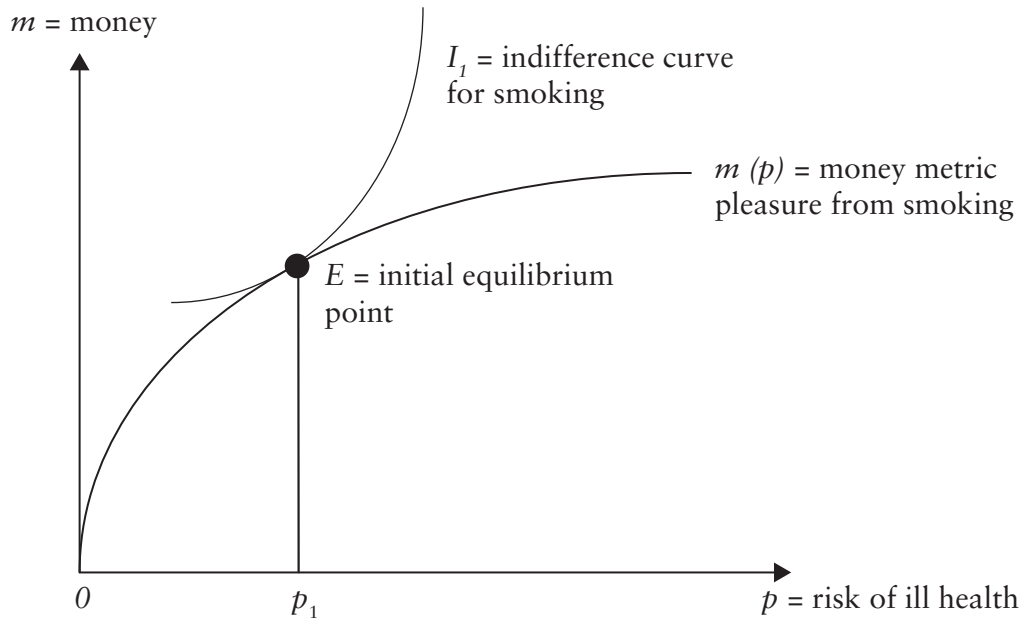


Figure 24.5 The Initial Equilibrium Where the Individual Chooses Point *E* Representing Constrained Utility Maximization

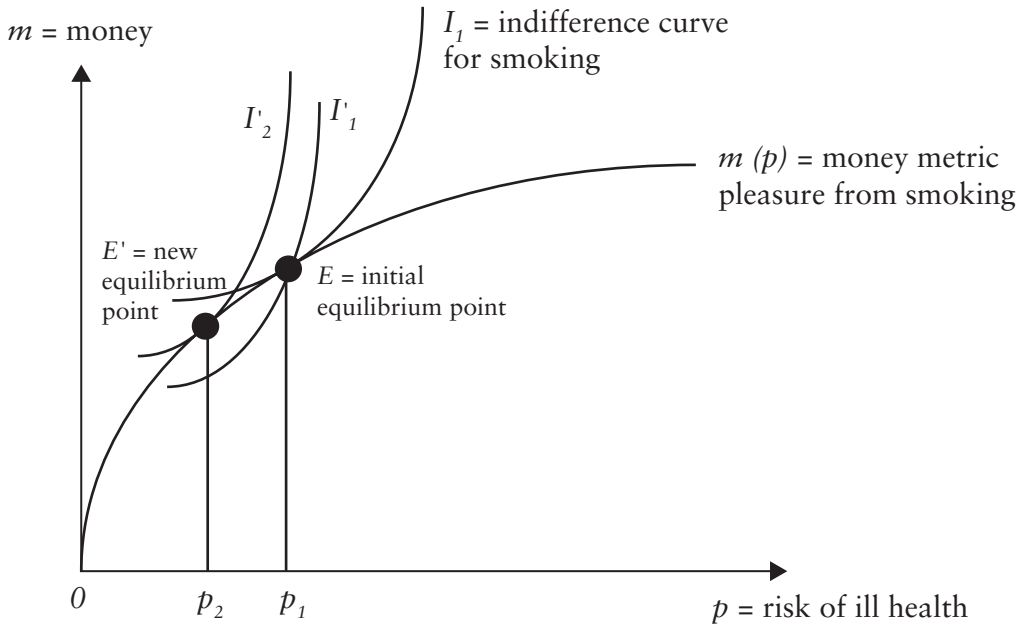


Figure 24.6 Increased Social Capital Causes the Individual to Re-Evaluate the Trade-Offs between Risk and Reward So That Lesser Risk Is Chosen

equilibrium at *E*. To be willing to add more risk he requires more money. Now Jack can improve to point *E´*, where he will smoke less. In fact, if Jack views *S* gains as a substitute for smoking, he necessarily will do this.

Could it happen instead that Jack views this gain in *S* as a complement to risk-taking? There is nothing wrong or illogical about that. The two may be looking forward to mountain climbing together, but here we are talking about cigarette smoking where it makes little sense.

Social Capital and Smoking

In addition to our assessment of risk-taking and social capital, other literature has suggested additional pathways that could lead to less smoking for individuals and for communities. Lindström (2009) notes that *S* could provide more trust in expert reports and help develop "political trust." Rocco (2014, p. 179) points out that social capital can enhance trust in the public and governmental institutions that usually take charge of antismoking campaigns.

Cigarette smoke has harmful and unpleasant external effects on people nearby, and our friends nearby influence us in turn.

> Smokers were also progressively found in the periphery of the social network. A spouse quitting decreased a person's chance of smoking by 76%. A sibling's quitting decreased the chance by 25%. A friend quitting decreased the chances by 36%. Among those quitting in small firms, a co-worker quitting decreased the chances by 34%.
>
> (Christakis, 2008, p.1)

Empirical support: Studies often find that subjects with strong social capital are less likely to smoke cigarettes and more likely to avoid risky drugs or excessive alcohol consumption. The most recent supporting studies are Takakura (2011); Tamubolon, Subramanian, and Kawachi (2011); and Rocco and d'Hombres (2014).

Conclusion

As we have noted, economic research on the health effects of social capital are quite new. They complement work by other social sciences as well as medicine and epidemiology by emphasizing models and statistical tests, not just correlations, but the exploration of causality issues by empirical means.

This chapter presents the representative results. The many studies which demonstrated a bivariate (negative) correlation between social capital and health have been supported by multivariate econometric studies. We have further explained the causal tests, and in two cases demonstrated a significant causal relationship; improvements in social capital tend to improve health—either individual or community health.

Other findings suggest fruitful avenues for future research. Trust, which is found to be beneficial to health, is also found to be higher within the more ethnically homogeneous areas. Perhaps these and other related studies would prove useful in addressing the world refugee situation. Also, the role of social capital in decreasing risk taking could be useful in decreasing the rate of drug addiction. Finally, modeling of social capital revealed possible pathways for metropolitan investments that would increase social capital and thus improve the average health of residents.

In summary, many results in this new, interdisciplinary field of study are certainly encouraging. They also have public health policy implications, as demonstrated by recent recommendations to upgrade the nation's approach to public health to emphasize cross-sector collaboration and to include actions that directly affect the social determinants of health (DeSalvo et al., 2016).

Summary

1 Social capital is defined as the extent of social networks, trust, and community norms.
2 Like other forms of capital, its level at a given point in time depends on the depreciation and investment.
3 Glaeser, Laibson, and Sacerdote explain the effect of changes in parameters (opportunity costs, average community social capital, the rate of payoff, the discount factor, the subject's age) on investment.
4 Community social capital may be increased by community efforts to lower the cost and safety of meeting people, and participation in the community.
5 Putnam's correlations revealed beneficial associations of social capital and measures of health by U.S. states.
6 Health economists develop multivariate regression studies that improve on the evidence for the Social Capital and Health Hypothesis.
7 A crucial test was done successfully and this further supports the hypothesis that the relationship is *causal*.
8 *Reverse causality* occurs as well.
9 In theory, one's inclination toward taking risks is reduced by his or her social capital.
10 Increases in social capital generally yield reductions in smoking.

Discussion Questions

1 If you were part of the City Council, what would you propose to increase community social capital?
2 From a cost/benefit standpoint, given that these are benefits to community social capital, what are the costs?
3 Why would an increase in average community social capital, $\hat{S}$, improve the payoff to the individual's social capital, S?
4 Do the significant correlations between $\hat{S}$ and per capita mortality rates prove the Social Capital and Health Hypothesis? Why or why not?
5 Reverse causality means that *H implies S*. Suppose this were the only result from these studies. How would community policy need to change from those based on *S implies H*?
6 Describe public health policies to reduce smoking levels in the community. What, if anything, would you choose?

Exercises

1 Consider the parameters R, $\hat{S}$, *S, w, and* $C(I)$ in equation (24.1). Find and explain the effect of an increase in each one.
2 For Table 24.1, suppose you think that you omitted a relevant variable, *physicians per capita*. Would this harm your results for $\hat{S}$ or H? Explain your view.
3 Suppose you moved into a neighborhood and that as time passes you found that your neighbors were not only friendly but dependable. How would you alter your risky behavior? Explain.

Note

1 Served on committees of a local organization; served as an officer of a club or organization; civic and social organizations per 1,000 population; mean number of club meetings attended last year; mean number of group memberships; turnout in presidential elections; attended public meetings; nonprofit organizations per 1,000 population; mean times worked on community projects last year; mean times did volunteer work last year; agree that "I spend a lot of times visiting friends"; mean times entertained at home last year; agree that "most people can be trusted"; agree that "most people are honest."

Glossary

Accountable Care Organizations (ACOs) Groups of doctors, hospitals, and other health care providers who organize voluntarily to give coordinated high-quality care to the populations they serve.

Actuarially Fair Insurance under which expected payouts equal the premiums paid by beneficiaries.

Adverse Selection A situation often resulting from asymmetric information in which individuals are able to purchase insurance at rates that are below actuarially fair rates plus loading costs.

Affordable Care Act (ACA) Signed into law by President Barack Obama in 2010 as the Patient Protection and Affordable Care Act. The legislation introduced numerous reforms but is best known for various provisions that have expanded health insurance coverage.

Agency Relationship A situation in which one person (agent) makes decisions on behalf of another person (principal).

Alternative Delivery System (ADS) Insurance and organizational arrangements for health care delivery that are alternatives to traditional fee-for-service (FFS) arrangements.

Asymmetric Information Situations in which the parties on the opposite sides of a transaction have differing amounts of relevant information.

Average Cost Total cost represents the sum of all fixed costs and variable costs in the short run. Average cost equals total cost divided by the quantity of output and also equals the sum of average variable cost (AVC) and average fixed cost (AFC). In the long run, average total cost represents the minimum possible cost per unit of producing any given level of output when there are no fixed costs.

Balance Billing The practice of collecting the difference between the charge and the insurance reimbursement from the patient.

Barriers to Entry Impediments to the unrestricted flow of factors into or out of an industry or occupation (e.g., control over natural resources, licensure, patents).

Body Mass Index (BMI) A measurement of tendency toward obesity, BMI = Weight in kilograms / (Height in meters)^2. Current usage defines a BMI over 25 as overweight, with BMI greater than 30 as obese.

Budget Constraint The line representing combinations of goods that the consumer is just able to afford, in a consumer optimization problem.

Capitation A method of reimbursement in managed care plans in which a provider is paid a fixed amount per person over a given period regardless of the amount of services rendered.

Cardinal Utility A quantitative measure of the value of a good in terms of metrically measurable utility. It is used in the study of risk and insurance.

Case-Mix Index A numerical measure of the intensity of patient cases treated by a given hospital, so that a higher value indicates a greater average degree of complexity of the cases.

Centers for Medicare & Medicaid Services (CMS) An agency of the U.S. Department of Health and Human Services (DHHS). CMS is responsible for administering the financing and quality assurance programs for Medicare and the federal participation in Medicaid; formerly the Health Care Financing Administration (HCFA).

Certificate-of-Need (CON) Regulations that require health care providers to obtain approval from state planning agencies for capital expenditures that exceed various threshold levels (e.g., $500,000).

Ceteris paribus Latin: Other things being held constant.

CHIP A program administered by CMS that provides matching funds to states for health insurances for families with children. The program was initially designed to cover uninsured children in families with incomes that are modest but too high to qualify for Medicaid. Formerly known as SCHIP.

Coefficient of Variation A measure of dispersion equal to the standard deviation divided by the mean (and sometimes multiplied by 100).

Coinsurance (Rate) The share of costs (fraction or percent) paid by the beneficiary of a health policy (often after some deductibles).

Community Rating The practice of setting insurance premiums based on the utilization pattern of a broad population in a region. This approach to rate setting contrasts with experience rating.

Comparative Statics The analysis that calculates the level of a new equilibrium given changed values of one or more economic parameters, such as prices or income.

Competition (*See* Perfect Competition)

Concentration Ratio The share of the market sales or production accounted for by a certain number of the largest firms. Often the four-firm ratio is used.

Consumer-Directed Health Plan (CDHP) A high-deductible health plan (HDHP) coupled with a tax-advantaged health spending account (HSA or HRA). Consumers are provided with information and tools to help with health care service and financing decisions.

Copayment An amount paid out-of-pocket by the insurance beneficiary as a result of coinsurance and deductibles.

Cost-Benefit Analysis (CBA) A method of comparing the monetary value of all benefits of a social project with all costs of that project.

Cost-Effectiveness Analysis (CEA) A method that tries to find the least-cost method of achieving a desired objective(s) associated with a social project.

Cost Sharing (*See also* Copayment) Methods of financing health care that require some direct payments for services by patients.

Cost Shifting The practice by suppliers of increasing charges from some payers to offset uncompensated care costs and lower net payments from other payers.

Cost-Utility Analysis (CUA) A form of cost-effectiveness analysis in which outcomes, such as quality-adjusted life-years (QALYs), reflect the quantity and quality of life.

Cross-(Price)-Elasticity of Demand (*See also* Elasticity) The percentage change in the quantity demanded of one good resulting from a 1 percent change in the price of another good.

Crowd-Out A response to the introduction or enhancement of public insurance, referring to the extent that those who have previously used private coverage now use public coverage.

CT Scan (Computerized Tomography) A cross-sectional rendering of the head and/or body making use of computer-processed combinations of many X-ray images taken from different angles.

Deductible The amount of health care charges for which a beneficiary is responsible before the insurer begins payment.

Demand Function The relationship between quantity demanded and price (and other independent variables, such as income and tastes). One could study individual demand as well as market demand.

Depreciation The change in the value of a good over time, relating to change in productivity due to deteriorating physical characteristics or technical obsolescence.

Diagnosis Related Groups (DRGs) A set of case types established under the prospective payment system (PPS) identifying patients with similar conditions and processes of care. As of 2016, there are 745 Medicare Severity Long-Term Care Diagnostic Related Groups (MS-DRGs). DRGs have been implemented in many other countries, as well as the United States.

Discount Rate The interest rate used when converting sums to be received at a future date to a present value.

Discounting The process of converting sums to be received at a future date to a present value.

Economic Profit (*See also* Monopoly Profit) The return over and above that which is necessary to keep the firm from exiting the market over the long run. These profits are also called above-normal profits, excess profits, and supranormal profits.

Economies of Scale Situation in which the long-run average costs of a firm decline as output increases.

Economies of Scope Situation in which a firm can jointly produce two or more goods more cheaply than under separate production of the goods.

Edgeworth Box (in Consumption) A diagram that shows all possible allocations of fixed amounts of goods and services between two people.

Efficiency (*See also* Pareto Efficiency) Technical efficiency occurs when the firm produces the maximum possible sustained output from a given set of inputs. This idea is distinguished from allocative efficiency—situations in which either inputs or outputs are put to their best possible uses in the economy so that no further gains in output or welfare are possible. Both allocative and technical efficiency are prerequisites for Pareto efficiency.

Elasticity The percentage change in a dependent variable (e.g., quantity demanded) resulting from a 1 percent change in an independent variable (e.g., price). Elasticities that exceed 1 in absolute value are considered elastic; elasticities less than 1 are inelastic.

Elasticity of Substitution (*See also* Elasticity) The percentage change in the capital–labor ratio resulting from a 1 percent change in relative factor prices of capital and labor.

Equilibrium Price (Quantity) The price (quantity) at which the quantity demanded and the quantity supplied are equal.

Evidence-Based Medicine A decision and reimbursement process that integrates clinical practice with the best available scientific evidence. Ideally, the evidence is based on rigorous research methods or systematic literature reviews.

Expected Value A measure used with a probability distribution of returns. The expected value is the sum of each probability multiplied by its corresponding return.

Experience Good A good for which evaluation is difficult prior to experience or purchase.

Experience Rating The practice of setting insurance premiums for an individual or group based on historical experience or risk associated with the individual or group.

Externality A case in which a consumer (producer) affects the utility (costs) of another consumer (producer) through actions that lie outside the price system.

Fee-for-Service (FFS) A method of payment under which the provider is paid for each procedure or service that is provided to a patient.

Fee Schedule A listing of fees by third-party payers showing the maximum amounts they will reimburse for specific services or procedures.

Firm Any entity that transforms inputs to some product or service that is sold in the marketplace.

First Fundamental Theorem of Welfare Economics The proposition that under specified conditions, competitive markets lead to Pareto efficient results.

Fixed Costs (TFC and AFC) Costs that do not vary with output. They are expressed either as total fixed cost (TFC) or average fixed cost (AFC).

Formulary A list of drugs developed by a managed care plan. Under a positive formulary, prescriptions on the list are covered. Under a negative formulary, prescriptions on the list are not covered.

Frontier Analysis A statistical analysis of producer efficiency that attempts to identify the best possible production practice and interprets inefficiency as a departure from the best possible production practice or frontier.

Game Theory A model that analyzes economic behavior as a series of strategic moves and countermoves by rival agents or players.

Gatekeeper The primary care provider who is responsible for coordinating a patient's care in a managed care plan. Often the gatekeeper must authorize referrals to specialists and nonemergency hospital admissions.

Gross Domestic Product (GDP) The market value of final goods and services produced within the borders of a country over a period of one year.

Group Insurance An insurance contract in which employees or members of a group are covered by a policy issued by their employer or group.

Health and Health Status (*See also* Morbidity Rate and Mortality Rate) The measures of the physical and emotional well-being of an individual or a defined population. Mortality and morbidity rates are often used to measure health status.

Health Care Goods and services used as inputs to produce health. Some analyses consider people's own time and knowledge used to maintain and promote health, in addition to conventional health care inputs.

Health Insurance Marketplaces or Health Exchanges State-based or federally-facilitated marketplaces established under the ACA where individuals can purchase health insurance.

Health Insurance Portability and Accountability Act (HIPAA) Federal legislation enacted in 1996 to protect the portability and continuity of health insurance coverage for workers who change or lose their jobs. The act requires hospitals, doctors, and insurance companies to share patient medical records and personal information on a wider basis to combat waste and fraud. The act also contains privacy provisions to protect the confidentiality of identifiable health data.

Health Maintenance Organization (HMO) A managed care plan that integrates financing and delivery of a comprehensive set of health care services to an enrolled population. HMOs may contract with or directly employ health care providers.

Health Plan Employer Data and Information Set (HEDIS) A set of standardized measures to evaluate health plan performance. HEDIS is used by the National Committee on Quality Assurance (NCQA) to accredit HMOs.

Health Reimbursement Account (HRA) (*See also* HSA) The HRA is similar to a health savings account (HSA), but it is completely controlled by the employer and does not have to be linked to a high-deductible health plan. It is used to pay for qualified medical expenses and can also be used to purchase health insurance.

Health Savings Account (HSA) Introduced in 2003 as part of the Medicare prescription drug benefit legislation, the HSA is a less-restrictive medical savings account (MSA), owned by the employee, and open to anyone enrolled in a high-deductible health plan (HDHP) and not already covered by public or private insurance.

Herfindahl-Hirschman Index (HHI) A measure of market concentration that incorporates the size distribution of firms. It is found by summing the squares of the market shares of each firm and varies from 0 (no concentration) to 1 (pure monopoly), or (when shares are in percent terms) from 0 to 10,000.

High-Deductible Health Plan (HDHP) A health insurance plan with a much higher deductible and a lower insurance premium than a traditional plan. An individual must have an HDHP to open a health savings account (HSA).

Human Capital A form of intangible capital that includes the skills and other knowledge that workers possess, or acquire through education, training, and health care that yields valuable productive services over time.

Income Effect (*See also* Substitution Effect) The effect on quantity demanded that results from the change in real income associated with a relative change in the price of the good or service under study.

Income Elasticity of Demand (*See also* Elasticity) The percentage change in quantity demanded resulting from a 1 percent change in income.

Indemnity Insurance Traditional health insurance often associated with fee-for-service payments, fee schedules, and which, unlike managed care, places few restrictions on choice of providers.

Independent Practice Association (IPA) A grouping of physicians in independent, solo, or small group practices who contract with a managed care organization to provide services to members.

Indifference Curve Geometric construction showing all combinations of goods that provide a constant level of satisfaction (utility) to the individual under study.

Inefficiency (*See* Efficiency and Pareto Efficiency)

Infant Mortality Rate The ratio of the number of deaths in infants age one year or less during a year divided by the number of live births during the year.

Inferior Good (*See also* Normal Good) A good or service for which demand decreases as income increases.

Internal Rate of Return The discount rate that will equate the time streams of costs and returns of an investment. It is a measure of the profitability of an investment.

Isoquant (Isoproduct Curve) All combinations of factors of production yielding a constant level of output.

Labor–Leisure Trade-Off In graphing time allocation, the line representing the combinations of leisure time and earnings from work that are possible for a person.

Law of Demand A statement of the inverse relationship between price and quantity demanded, all else equal.

Law of Diminishing Returns After some point, the marginal product of a variable input must diminish.

Loading Costs The administrative and other costs associated with underwriting an insurance policy.

Long Run (*See also* Short Run) A period of time sufficient to permit a firm to vary all factors of production.

Long-Term Care The ongoing health and social services provided for individuals who need assistance on a continuing basis because of physical or mental disability. Services can be provided in an institution, the home, or the community, and include informal services provided by family or friends as well as formal services provided by professionals or agencies.

Luxury Good A good that richer people tend to buy in greater proportions so that its income elasticity exceeds +1.00.

Managed Care Any payment or delivery arrangement used by a health plan or provider to control or to coordinate use of health services to contain health expenditures, improve quality, or both.

Managed Care Organization (MCO) A general term referring to the various health plans that use managed care arrangements and have a defined system of selected providers that contract with them. The most common are health maintenance organizations (HMOs), preferred provider organizations (PPOs), and point-of-service (POS) plans.

Mandated Benefits The coverage in health insurance policies for services that are mandated by state insurance statutes.

Marginal Cost The increase in total cost resulting from a one-unit increase in output.

Marginal Labor (Factor) Cost The addition to total labor (factor) costs associated with an additional unit of labor (factor of production).

Marginal Product The addition to total output resulting from an additional unit of the variable input.

Marginal Rate of Substitution The amount of one commodity given up per unit increase in another commodity, while maintaining the same level of production or satisfaction (for consumers).

Marginal Rate of Technical Substitution The amount of one factor of production given up per unit increase in another factor of production, while maintaining the same level of output.

Marginal Rate of Transformation The slope of the production possibilities curve, and the rate at which society can transform one good into another.

Marginal Revenue The addition to total revenue associated with a one-unit increase in output.

Marginal Revenue Product The addition to a firm's total revenue associated with employing one more unit of a variable input.

Marginal Utility The extra utility gained from consuming one more unit of a good, holding others constant. Utility is a measure of the satisfaction from consuming goods.

Market Demand The total demand for a good by all consumers in the market.

Market Structure The organization of an industry in terms of the number and distribution of firms and how firms compete among themselves.

Medicaid The health insurance programs administered by the states for qualifying low-income beneficiaries. The federal government establishes minimum standards and provides matching grants. The program became law in 1965.

Medical Savings Account (MSA) A limited health spending account (HSA) that was introduced on an experimental basis in the 1990s. MSAs were aimed at small businesses and individuals.

Medicare The federal health insurance program established in 1965 for the elderly and other selected groups.

Medicare—Part A The Medicare Hospital Insurance that covers beneficiaries for inpatient hospital, home health, hospice, and limited skilled nursing facility services. Beneficiaries are responsible for deductibles and copayments.

Medicare—Part B The Medicare Supplementary Medical Insurance that covers Medicare beneficiaries for physician services, medical supplies, and other outpatient treatment. Beneficiaries are responsible for monthly premiums, copayments, deductibles, and balance billing.

Medicare—Part C (Medicare Advantage) An expanded set of options for the delivery of health care under Medicare. While all Medicare beneficiaries can receive their benefits through the original fee-for-service program, most beneficiaries enrolled in both Part A and Part B can choose to participate in a Medicare Advantage plan instead. Managed care or fee-for-service providers that seek to contract as Medicare Advantage plans must meet specific organizational, financial, and other requirements.

Medicare—Part D Part D provides subsidized access to prescription drug insurance coverage on a voluntary basis, upon payment of a premium, to individuals entitled to Part A or enrolled in Part B, with premium and cost-sharing subsidies for low-income enrollees. Beneficiaries may enroll in either a stand-alone prescription drug plan (PDP) or an integrated Medicare Advantage plan that offers Part D coverage.

Medigap Policy A privately purchased insurance policy that supplements Medicare coverage and meets specified requirements set by federal statute and the National Association of Insurance Commissioners.

Monopoly Situations in which a producer faces a negatively sloped demand curve. In a pure monopoly, no other firm produces a close substitute for the firm's product. The demand curve facing the monopolist is the market demand curve.

Monopoly Profit (Rent) The return over and above a normal profit resulting from monopoly.

Monopsony Situations in which a producer faces a positively sloped supply curve in the product or factor market because it is the only buyer. The supply curve facing the monopsonist is the market supply curve.

Moral Hazard A term that represents the disincentives created by insurance (more generally, any contractual arrangement) for individuals to take measures that would reduce the amount of care demanded. In the health services literature, it more commonly describes the additional quantity of health care demanded, due to a decrease in the net price of care attributable to insurance.

Morbidity Rate The rate of incidence of disease in a particular population.

Mortality Rate The death rate for a particular population. The crude death rate is the ratio of deaths during a year divided by midyear population. Because age is so important, the age-adjusted mortality rate is a measure that takes into account a population's age distribution.

Necessity A good whose consumption does not vary greatly with changes in peoples' incomes. More generally, a good with an income elasticity less than +1.00.

Nominal Value (*See also* Real Value) The money value measured in current dollars.

Nonprofit Firm A firm that is constrained by law from distributing any residual of income over costs to any party.

Normal Good (*See also* Inferior Good) A good or service for which demand increases as income increases.

Normal Return (Normal Profit) The return just sufficient to retain factors of production in an industry or an occupation in the long run. The return equal to the opportunity cost of a factor of production.

Nursing Facility An institution that provides skilled nursing care and rehabilitation services to injured, functionally disabled, or sick persons.

Offer Curve A set of points summarizing the amount of a good that an individual will offer for trade, given his or her preferences, endowment, and the prices of other goods.

Opportunity Cost The value of the best forgone alternative when consuming or producing more of the commodity under consideration.

Ordinal Utility Utility as evaluated through relative levels of satisfaction, when the particular unit of utility is not essential. Examples of ordinal numbers are first, second, and third.

Pareto Efficiency (Optimum) (*See also* Efficiency) An allocation in which it is impossible to improve the level of welfare of one party without hurting the welfare level of another party. Circumstances in which the level of welfare of one or more parties can be improved without hurting any other party are Pareto improvements.

Participation/Assignment A situation in which a provider agrees to accept the third-party payer's payment in full, thereby relieving the patient of any balance (except for applicable patient copayments).

Pay-for-Performance (P4P) The effort by managed care organizations and other payers to reward providers who improve the quality of their care by meeting certain performance standards. Providers typically receive bonus payments for meeting the goals.

Peer Review Organization (PRO) An organization that contracts with the Centers for Medicare & Medicaid Services to investigate the quality of health care furnished to Medicare beneficiaries and to educate beneficiaries and providers. PROs also conduct limited review of medical records and claims to evaluate the appropriateness of care provided.

Perfect Competition A market structure with (1) numerous buyers and sellers, (2) perfect information, (3) free entry and exit, and (4) a homogeneous product.

Play or Pay Insurance proposal that would require employers to either provide minimal levels of health insurance to their employees ("play") or *pay* into a fund that would be used to provide coverage.

Point-of-Service (POS) Plan A managed care plan that encourages patients to select a provider in a network. Members can select non-network providers but will incur higher out-of-pocket costs.

Preferred Provider Organization (PPO) An arrangement under which an enrollee is given financial incentives (e.g., zero copayments) to seek care from selected physicians and hospitals with which the payer has contracted.

Prepaid Group Practice (PGP) A prepayment arrangement under which participating providers agree to provide services to eligible enrollees in return for a fixed capitated payment.

Prepayment (Prepaid Plans) Health insurance proposal that provides unlimited amounts of covered services in return for a fixed predetermined premium.

Present (Discounted) Value (PV or PDV) (*See also* Discounting) The value of a stream of returns to be received at future dates, which is discounted to the equivalent of present dollars.

Prevalence (*See also* Incidence) In epidemiology, the fraction of the population that is currently infected. Incidence adds new cases to the total pool describing the prevalence of present cases.

Price Discrimination The sale of goods or services to different individuals at different prices.

Price Elasticity of Demand (*See also* Elasticity) The percentage change in quantity demanded resulting from a 1 percent change in price.

Price Elasticity of Supply (*See also* Elasticity) The percentage change in quantity supplied resulting from a 1 percent change in price.

Price Index A measure of the current prices of a bundle of goods relative to the prices of these goods in a base year. A price index, often used to convert nominal values to real values, shows how much the price of that bundle has changed since the base year.

Production Function The relationship between the maximum output that can be produced corresponding to any combination of factor inputs.

Production Possibilities Curve (Transformation Curve) A relationship showing all combinations of goods that an economy can produce with given amounts of input factors and the existing technology. The slope of the curve is the marginal rate of transformation showing the amount of one good that must be given up for a one-unit increase in the other good.

Prospective Payment System (PPS) The method of hospital reimbursement phased in by Medicare beginning in 1983 under which hospitals were reimbursed a fixed amount determined by the diagnostic related groups of their admissions.

Public Good (Pure) A good (e.g., national defense) that no one can be prevented from consuming (i.e., nonexcludable) and that can be consumed by one person without depleting it for another (i.e., nonrival). The marginal cost of providing the good to another consumer is zero.

Quality-Adjusted Life-Year (QALY) A measure of health outcome that incorporates quantity and quality of life. It uses a weighting system that assigns a value ranging from 1 (perfect health) to 0 (health state equivalent to death).

Real Value (*See also* Nominal Value) Monetary value that is adjusted for changes in the general level of prices relative to some arbitrarily selected base year.

Regression Analysis Statistical analysis that posits a linear relationship between the variable to be explained y and one or more (in multiple regression) explanatory variables x, in the form $y = a + bx$.

Reinforcement A characteristic of addictive behavior indicating that greater past consumption of addictive goods increases the desire for present consumption.

Rent (Economic Rent) The remuneration to a factor of production, over and above the amount that is necessary to induce its supply in the market.

Reputation Good A good for which consumers rely on information provided by friends, neighbors, and others.

Risk Aversion The degree to which a certain income or wealth is preferred to a risky alternative with the same expected income or wealth.

Risk Selection The enrollment choices made by health plans or enrollees on the basis of perceived risk relative to the premium to be paid.

Search Good A good whose characteristics can be fully evaluated upon inspection.

Second Fundamental Theorem of Welfare Economics The proposition that any Pareto efficient outcome can, in principle, be achieved by competitive markets, given an appropriate initial endowment.

Selective Contracting The practice of a managed care organization (MCO) by which the MCO enters into participation agreements only with certain providers (and not with all providers who qualify) to provide health care services to health plan participants as members of the MCO's provider panel.

Short Run (*See also* Long Run) Situations in which the firm is not able to vary all its inputs. There is at least one factor of production that is fixed.

Single-Payer System Payment to health care providers administered by a single entity or organization, usually, but not necessarily, the government. Canada is often cited as an example of a single-payer system.

Small Area Variations The large variations in the per capita rates of utilization across small, homogeneous areas for many medical and surgical procedures.

Social Insurance Government insurance programs in which eligibility and premiums are not determined by the practices common to private insurance contracts. Premiums are often subsidized and there are typically redistributions from some segments of the population to others.

Social Welfare Function A decision rule under which a society ranks all possible distributions of goods and services.

Staff Model HMO A health maintenance organization (HMO) in which physicians are directly employed by the HMO.

Substitutes Substitutes in consumption are goods that satisfy the same wants (e.g., beef and chicken) so that an increase in the price of one will increase the demand for the other. Substitutes in production are alternative goods that a firm can produce (e.g., corn and soybeans for a farmer) so that an increase in the price of one will lead to a decrease in the supply of another.

Substitution Effect (*See also* Income Effect) The change in quantity demanded resulting from a relative change in commodity prices, holding real income constant.

Supplier-Induced Demand (SID) The change in demand associated with the discretionary influence of providers, especially physicians, over their patients. Demand provided for the self-interests of providers rather than solely for patient interests.

Sustainable Growth Rate (SGR) A formula, repealed in 2015, under which Medicare's physician fees were tied to a target based on caseloads, practice costs, and the gross domestic product. When current spending on physician services exceeded the targets, the formula called for fee cuts to be applied prospectively.

Take-Up A response to the introduction or enhancement of public insurance, referring to the extent that those who have previously been uninsured now use public coverage.

Target Income Hypothesis A model under which providers are thought to select a specified income level, and to adjust their amount of services provided or fees in order to reach this level.

Technical Efficiency (*See* Efficiency)

Technological Change A change in the process by which factors of production combine to produce outputs.

Theorem of the Second Best The economic theorem stating that the correction of some but not all market imperfections, in cases where there is more than one imperfection, may not necessarily improve society's welfare.

Time Costs The money value of the time lost through travel or waiting when consuming a product or service.

Tolerance A characteristic of addictive behavior indicating that the incremental utility from a given amount of consumption of the addictive good is lower when past consumption is greater. This suggests that the marginal future impacts of smoking or drinking or ingesting drugs decrease when we consume more at the present time.

Uncompensated Care The care rendered by hospitals or other providers without payment from the patient or a government-sponsored private insurance program. It includes both charity care, provided without expectation of payment, and bad debts, for which the provider has made an unsuccessful effort to collect payment due from the patient.

Utility and Utility Function Utility represents satisfaction or the level of welfare of an individual, measured in cardinal or ordinal utility terms. The utility function expresses the person's utility as a function of all possible combinations of goods and services.

Utilization Review (UR) The programs that attempt to determine whether specific services are medically necessary and delivered at an appropriate level and cost.

Value-Based Insurance Design (VBID) An insurance program design that reduces patients' out-of-pocket costs for demonstrated high-value services. The services typically involve primary preventive care.

Variable Costs (TVC and AVC) Costs associated with variable factor(s) of production, often expressed as total variable cost (TVC) or average variable cost (AVC).

Welfare A measure of an individual's or a society's level of well-being.

Welfare Loss or Deadweight Loss A measure of the net loss of society's welfare resulting from a misallocation of resources, usually situations in which the marginal benefits of a good do not equal marginal costs.

Yardstick Competition A regulatory pricing policy in which an average of the marginal costs of all competing firms is used as a standard of payment to induce the firm to engage in cost-cutting innovation.

References

Aaberge, Rolf et al., "The Distributional Impact of Public Services When Needs Differ," *Journal of Public Economics 94* (2010): 549–562.

Aaron, Henry J., "The Costs of Health Care Administration in the United States and Canada—Questionable Answers to a Questionable Question," *New England Journal of Medicine 349* (2003): 801–803.

Abbott, Thomas A., "Price Regulation in the Pharmaceutical Industry: Prescription or Placebo?" *Journal of Health Economics 14* (1995): 551–565.

Absolo, Ignacio, and Aki Tsuchiya, "Exploring Social Welfare Functions, and Violations of Monotonicity: An Example from Inequalities in Health," *Journal of Health Economics 23* (2004): 313–329.

Acemoglu, Daron, and Amy Finkelstein, "Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector," *NBER* Working Paper w12254, 2006.

Acemoglu, Daron, Amy Finkelstein, and Matthew J. Notowidigdo, "Income and Health Spending: Evidence from Oil Price Shocks," *Review of Economics and Statistics 95* (2013): 1079–1095.

Acton, Jan P., "Demand for Health Care among the Urban Poor with Special Emphasis on the Role of Time," in *The Role of Health Insurance in the Health Services Sector*, Richard Rosett (ed.) (New York: Neal Watson, 1976).

Acton, Jan P., "Nonmonetary Factors in the Demand for Medical Services: Some Empirical Evidence," *Journal of Political Economy 83* (1975): 595–614.

Adams, Scott J., "Health Insurance Market Reform and Employee Compensation: The Case of Pure Community Rating in New York," *Journal of Public Economics 91* (2007): 1119–1133.

Agency for Healthcare Research and Quality, *2014 National Healthcare Quality and Disparities Report*, Washington, 2015.

Agha, Leila, "The Effects of Health Information Technology on the Cost and Quality of Medical Care," *Journal of Health Economics 34* (2014): 19–30.

Aigner, Dennis J., C. A. Knox Lovell, and Peter Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics 6* (1977): 21–37.

Aizer, Anna, Janet Currie, and Enrico Moretti, "Does Managed Care Hurt Health? Evidence from Medicaid Mothers," *Review of Economics and Statistics 89* (2007): 385–399.

Akerlof, George A., "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics 84* (1970): 488–500.

Alexander, Jeffrey A., et al., "Do Market-Level Hospital and Physician Resources Affect Small Area Variation in Hospital Use?" *Medical Care Research and Review 56* (1999): 94–117.

Allin, Sara, and David Rudollir, "The Canadian Health Care System, 2015," in Elias Mossialos, et al., *2015 International Profiles of Health Care Systems*, The Commonwealth Fund, January 2016, X-1857_Mossialos_intl_profiles_2015_v7.pdf, www.commonwealthfund.org/~/media/files/publications/fund-report/2016/jan/1857_mossialos_intl_profiles_2015_v7.pdf, last accessed June 16, 2016.

Almond, Douglas, et al., "Estimating Marginal Returns to Medical Care: Evidence from at-Risk Newborns," *Quarterly Journal of Economics 125* (2010): 591–634.

Altman, Stuart H., David Shactman, and Eilat Efrat, "Could U.S. Hospitals Go the Way of U.S. Airlines?" *Health Affairs 25* (2006): 11–21.

Alvarez-Dardet, Carlos, and Maria Teresa Ruiz, "Thomas McKeown and Archibald Cochrane: A Journey through the Diffusion of Their Ideas," *British Medical Journal 306* (1993): 1252–1254.

Alviola, Pedro A., R.M. Nyga, M.R. Thomsen, et al., "The Effect of Fast Food Restaurants on Childhood Obesity: A School Level Analysis," *Economics and Human Biology 12* (2014): 110–119.

American Medical Association, *Competition in Health Insurance: A Comprehensive Study of U.S. Markets 2010 Update* (AMA: Chicago, 2010).

Amin, V., Berhrman, J., Spector, T., "Does More Schooling Improve Health Outcomes and Health Related Behaviors?" *Economics of Education Review 35* (2013): 134–148.

Andermann, Anne Adina Judith, "Physicians, Fads, and Pharmaceuticals: A History of Aspirin," *McGill Journal of Medicine 2* (1996), http://archive.li/crtO, accessed July 8, 2016.

Anderson, Gerard F., and Jean-Pierre Poullier, "Health Spending, Access, and Outcomes, Trends in Industrialized Countries," *Health Affairs 18* (1999): 178–192.

Anderson, Gerard F., Peter S. Hussey, Bianca K. Frogner, and Hugh R. Waters, "Health Spending in the United States and the Rest of the Industrialized World," *Health Affairs 24* (2005): 903–914.

Anderson, Gerard F., Uwe E. Reinhardt, Peter S. Hussey, and Varduhi Petrosyan, "It's the Prices, Stupid: Why the United States Is So Different from Other Countries," *Health Affairs 22* (2003): 89–105.

Anderson, Lisa R., Mellor, Jennifer, and Milyo, Jeffrey, "Social Capital and Contributions in a Public-Goods Experiment," *American Economic Review 94* (2004): 373–376.

Anderson, Odin W., *Health Services as a Growth Enterprise in the United States Since 1875*, 2nd ed. (Ann Arbor, MI: Health Administration Press, 1990).

Anderson, Patricia M., Kristin F. Butcher, and Phillip B. Levine, "Maternal Employment and Overweight Children," *Journal of Health Economics 22* (2003): 477–504.

Andreoni, James, "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal 100* (1990): 464–477.

Archibald, Robert, Clyde Haulman, and Carlisle Moody, "Quality, Price, Advertising, and Published Quality Ratings," *Journal of Consumer Research 9* (1983): 347–356.

Arnould, Richard J., Lawrence W. Debrock, and John W. Pollard, "Do HMOs Produce Specific Services More Efficiently?" *Inquiry 21* (1984): 243–253.

Arrow, Kenneth J., "Gifts and Exchanges," *Philosophy & Public Affairs 1* (1972): 343–362.

Arrow, Kenneth J., *Social Choice and Individual Values* (New York: Wiley, 1975).

Arrow, Kenneth J., "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice," *Journal of Philosophy 70* (1973): 245–262.

Arrow, Kenneth J., "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review 53* (1963): 941–973.

Arterburn, David, et al., "Introducing Decision Aids at Group Health Was Linked to Sharply Lower Hip and Knee Surgery Rates and Costs," *Health Affairs* 31 (2012): 2094–2104.

Atkinson, Scott E., and Thomas Crocker, "Econometric Health Production Functions: Relative Bias from Omitted Variables and Measurement Error," *Journal of Environmental Economics and Management 22* (1992): 12–24.

Auerbach, David I., et al., "Nurse-Managed Health Centers and Patient-Centered Medical Homes Could Mitigate Expected Primary Care Physician Shortage," *Health Affairs 32* (2013): 1933–1941.

Auld, M. Christopher, and Nirmal Sidhu, "Schooling, Cognitive Ability and Health," *Health Economics 14* (2005): 1019–1034.

Austin, D. Andrew, and Thomas L. Hungerford, "The Market Structure of the Health Insurance Industry," Congressional Research Service, 7–5700, December 17, 2010.

Ayanian, John Z., "Michigan's Approach to Medicaid Expansion and Reform," *New England Journal of Medicine 369* (2013): 1773–1775.

Aylward, Bruce, and Tadataka Yamada, "The Polio Endgame," *New England Journal of Medicine 364* (June 15, 2011): 2273–2275.

Bae, Jay P., "Drug Patent Expiration and the Speed of Generic Entry," *Health Services Research 22* (1997): 87–101.

Baicker, Katherine, et al., "The Oregon Experiment—Effects of Medicaid on Clinical Outcomes," *New England Journal of Medicine 368* (2013): 1713–1722.

Baicker, Katherine, Michael E. Chernew, and Jacob Robbins, "The Spillover Effects of Medicare Managed Care: Medicare Advantage and Hospitalization," *Journal of Health Economics 32* (2013): 1289–1300.

Bain, Joe S., *Barriers to New Competition* (Cambridge, MA: Harvard, 1956).

Bajari, Patrick, et al., "Moral Hazard, Adverse Selection, and Health Expenditures," *Rand Journal of Economics 45* (2014): 747–763.

Baker, Laurence C., "Managed Care and Technology Adoption in Health Care: Evidence from Magnetic Resonance Imaging," *Journal of Health Economics 20* (2001): 395–421.

Baker, Laurence C., and Kenneth S. Corts, "The Effects of HMOs on Conventional Insurance Premiums: Theory and Evidence," *NBER* Working Paper 5356, 1995.

Baker, Laurence C., and Kenneth S. Corts, "HMO Market Penetration and the Cost of Health Care: Market Discipline or Market Segmentation," *American Economic Review 86* (1996): 389–394.

Baker, Lawrence, and Ciaran Phibbs, "Managed Care, Technology Adoption, and Health Care: The Adoption of Neonatal Intensive Care," *NBER* Working Paper w7883, 2000.

Baker, Laurence C., and Joanne Spetz, "Managed Care and Medical Technology Growth," in *Frontiers in Health Policy Research, Volume 2*, Alan Garber (ed.) (Cambridge, MA: MIT Press, 2000).

Bala, Mohan V., and Gary A. Zarkin, "Are QALYs an Appropriate Measure for Valuing Morbidity in Acute Diseases?" *Health Economics 9* (2000): 177–180.

Balcombe, Kelvin, Ian Fraser, and Salvatore Di Falco, "Traffic Lights and Food Choice: A Choice Experiment Examining the Relationship between Nutritional Food Labels and Price," *Food Policy 35* (2010): 211–220.

Balia, Silvia, and Andrew Jones, "Mortality, Lifestyle and Socio-Economic Status," *Journal of Health Economics 27* (2008): 1–26.

Ballou, Jeffrey P., "Do the Nonprofit and Government Homes Enter Unprofitable Markets?" *Economic Inquiry 46* (2008): 241–260.

Ballou, Jeffrey P., and Burton A. Weisbrod, "Managerial Rewards and the Behavior of For-Profit, Governmental, and Nonprofit Organizations: Evidence from the Hospital Industry," *Journal of Public Economics 87* (2003): 1895–1920.

Balsa, Ana I., Zhun Cau, and Thomas G. McGuire, "Does Managed Care Reduce Health Care Disparities between Minorities and Whites?" *Journal of Health Economics 26* (2007): 101–121.

Baltagi, Badi H., and James M. Griffin, "The Econometrics of Rational Addiction: The Case of Cigarettes," *Journal of Business and Economic Statistics 19* (2001): 449–454.

Baltagi, Badi H., and Dan Levin, "Estimating Dynamic Demand for Cigarettes Using Panel Data: The Effects of Bootlegging, Taxation, and Advertising Reconsidered," *Review of Economics and Statistics 68* (1986): 148–155.

Banerjee, Rajabrata, "Competition and the Cost of Health Care," *Applied Economics 44* (2012): 1201–1207.

Barbetta, Gian Paolo, Gilberto Turati, and Angelo M. Zago, "Behavioral Differences between Public and Private Not-for-Profit Hospitals in the Italian National Health Service," *Health Economics 16*: 75–96 (2007).

Barry, Michael J., et al., "Patient Reactions to a Program Designed to Facilitate Patient Participation in Treatment Decisions for Benign Prostatic Hyperplasia," *Medical Care 33* (1995): 771–782.

Bates, Laurie J., and Rexford E. Santerre, "Do Health Insurers Posses Monopsony Power in the Hospital Service Industry?" *International Journal of Health Care Finance and Economics 8* (2008): 1–11.

Baumol, William J., John C. Panzar, and Robert Willig, *Contestable Markets and the Theory of Industry* (San Diego: Harcourt Brace Jovanovich, 1982).

Bays, Carson W., "Why Most Private Hospitals Are Nonprofit," *Journal of Policy Analysis and Management 2* (1983): 366–385.

Becker, Gary S., and Kevin M. Murphy, "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics 108* (1993): 941–964.

Becker, Gary S., and Kevin M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy 96* (1988): 675–700.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy, "Rational Addiction and the Price on Consumption," *American Economic Review 81* (1991): 237–241.

Becker, Gary S., Kevin M. Murphy, and Michael Grossman, "The Economic Theory of Illegal Goods: The Case of Drugs," *NBER* Working Paper 10976, 2004.

Becker, Gary S., Kevin M. Murphy, and Tomas Philipson, "The Value of Life Near Its End and Terminal Care," *NBER* Working Paper 13333, 2007.

Benham, Lee, "The Effects of Advertising on the Price of Eyeglasses," *Journal of Law and Economics 15* (1972): 421–477.

Berg, Bernard van den, Denzil Fiebig, and Jane Hall, "Well-Being Losses Due to Care-Giving," *Journal of Health Economics 35* (2014): 123–131.

Bhattacharya, Jayanta, "Specialty Selection and Lifetime Returns to Specialization within Medicine," *Journal of Human Resources 60* (2005): 115–143.

Bhattacharya, Jay, and M. Kate Bundorf, "The Incidence of the Healthcare Costs of Obesity," *NBER* Working Paper 11303, 2005.

Bhattacharya, Jay, and Darius N. Lakdawalla, "The Labor Market Value of Health Improvements," *Forum for Health Economics & Policy 9* (2006): (Biomedical Research and the Economy), Article 2. www.bepress.com/fhep/biomedical_research/2, accessed May 20, 2008.

Bilodeau, David, Pierre-Yves Cremieux, and Pierre Oullette, "Hospital Technology in a Non-market Health Care System," *Southern Economic Journal 68* (2002): 511–529.

Bishop, Christine E., "Where Are the Missing Elders? The Decline in Nursing Home Use, 1985 and 1995," *Health Affairs 18* (1999): 146–155.

Blank, Rebecca M., "Can Equity and Efficiency Complement Each Other?" *Labour Economics 9* (2002): 451–468.

Bleichrodt, Han, and John Quiggin, "QALYs and Consumer Demand for Health Care," *Journal of Health Economics 18* (1999): 681–708.

Blendon, Robert J., Mollyann Brodie, and John Benson, "What Happened to Support for the Clinton Plan?" *Health Affairs 14* (1995): 7–23.

Blendon, Robert, et al., "Common Concerns amid Diverse Systems: Health Care Experiences in Five Countries," *Health Affairs 22* (2003): 106–121.

Blendon, Robert J., et al., "Understanding the Managed Care Backlash," *Health Affairs 17* (1998): 80–94.

Blomquist, Ake, "QALYs, Standard Gambles, and the Expected Budget Constraint," *Journal of Health Economics 21* (2002): 181–195.

Blomquist, Ake G., and R. A. L. Carter, "Is Health Care Really a Luxury?" *Journal of Health Economics 16* (1997): 207–230.

Blomquist, Ake, and Pierre Thomas Leger, "Information Asymmetry, Insurance, and the Decision to Hospitalize," *Journal of Health Economics 24* (2005): 775–793.

Blumberg, Linda J., and Shanna Rifkin, "Stabilizing Premiums under the Affordable Care Act: State Efforts to Reduce Adverse Selection," *Urban Institute*, November 2013.

Blumberg, Linda J., John Holahan, and Erik Wengle, *Increases in 2016 Marketplace Nongroup Premiums: There Is No Meaningful National Average*, Robert Wood Johnson Foundation and Urban Institute, May 2016.

Blundell, R., and F. Windmeijer, "Identifying Demand for Health Resources Using Waiting Time Information," *Health Economics 9* (2000): 465–474.

Bodenheimer, Thomas S., and Mark D. Smith, "Primary Care: Proposed Solutions to the Physician Shortage without Training More Physicians," *Health Affairs 32* (2013): 1881–1886.

Böhm, Katharina, et al., "Five Types of OECD Healthcare Systems: Empirical Results of a Deductive Classification," *Health Policy 113* (2013): 258–269.

Bolin, Kristian, Lena Jacobson, and Bjorn Lindgren, "The Family as the Health Producer—When Spouses Act Strategically," *Journal of Health Economics 21* (2002): 475–495.

Bonsang, Eric, "Does Informal Care from Children to Their Elderly Parents Substitute for Formal Care in Europe?" *Journal of Health Economics 28* (2009): 143–154.

Borisova, Natalia, and Allen C. Goodman, "The Effects of Time and Money Prices on Treatment Attendance for Methadone Maintenance Clients," *Journal of Substance Abuse Treatment 26* (2004): 345–352.

Borjas, George J., H. E. Frech III, and Paul B. Ginsburg, "Property Rights and Wages: The Case of Nursing Homes," *Journal of Human Resources 17* (1983): 232–246.

Bosmans, Nathalie, and Fabienne Fecher, "Performance of Belgian Hospitals: A Frontier Approach," *Health Economics 4* (1995): 389–397.

Bosworth, Barry, and Gary Burtless, "Effects of Tax Reform on Labor Supply, Investment, and Saving," *Journal of Economic Perspectives 6* (1992): 3–25.

Boukas, Ellyn R., Alwyn Cassil, and Ann S. O'Malley, "A Snapshot of US Physicians: Key Findings from the 2008 Health Tracking Physician Survey," Data Bulletin No. 35, Center for Studying Health System Change, September 2009.

Bourdieu, P., *The Form of Capital* (New York: Greenwood, 1985).

Brekke, Kurt R., et al., "Competition and Quality in Health Care Markets: A Differential-Game Approach," *Journal of Health Economics 29* (2010): 508–523.

Brennan, Troyen A., et al., "Incidence of Adverse Events and Negligence in Hospitalized Patients," *New England Journal of Medicine 324* (1991): 370–376.

Brickley, James A., and R. Lawrence Van Horn, "Managerial Incentives in Nonprofit Organizations: Evidence from Hospitals," *Journal of Law and Economics 45* (2002): 227–249.

Brody, Howard, "Medicine's Ethical Responsibilities for Health Care Reform—The Top Five List," *New England Journal of Medicine 362* (2010): 283–285.

Brouwer, Werner B. F., "Need for Differential Discounting of Costs and Health Effects in Cost Effectiveness Analysis," *British Medical Journal 331* (2005): 446–448.

Brouwer, Werner B. F., et al., "Welfarism vs. Extra-Welfarism," *Journal of Health Economics 27* (2008): 325–338.

Brown, Douglas M., "Do Physicians Underutilize Aides?" *Journal of Human Resources 23* (1988): 342–355.

Brown, Lawrence D., "The National Politics of Oregon's Rationing Plan," *Health Affairs 10* (1991): 28–51.

Brown, Timothy T., Colla, C. H., and Scheffler, Richard M., "Does Community Level Social Capital Affect Smoking Behavior? An Instrumental Variables Approach," *Working Paper*, 2010, University of California: Berkeley.

Brown, Timothy T., et al., "The Empirical Relationship between Community Social Capital and the Demand for Cigarettes," *Health Economics 15* (2006): 1159–1172.

Browne, Mark J., and Helen I. Doerpinghaus, "Information Asymmetries and Adverse Selection in the Market for Individual Medical Expense Insurance," *Journal of Risk and Insurance 60* (1993): 300–312.

Brunello, Giorgio and Béatrice D'Hombres, "Does Body Weight Affect Wages? Evidence from Europe," *Economics and Human Biology 5* (2007): 1–19.

Bryan, Stirling, and Sue Jowett, "Hypothetical versus Real Preferences: Results from an Opportunistic Field Experiment," *Health Economics, 19* (2010): 1502–1509.

Buchanan, James, *The Limits of Liberty* (Chicago: University of Chicago Press, 1975).

Bundorf, M. Kate, et al., "Do Markets Respond to Quality Information? The Case of Fertility Clinics," *Journal of Health Economics 28* (2009): 718–727.

Burgess, James F. Jr., and Paul W. Wilson, "Variation in Inefficiency in U.S. Hospitals," *INFOR 36* (1998): 84–102.

Burke, Mary, Gary Fournier, and Kislaya Prasad, "Symposium: Third Annual Southwestern Health Economics Study Group Conference: The Diffusion of a Medical Innovation: Is Success in the Stars?" *Southern Economic Journal 73* (2007): 588–603.

Burns, Lawton R., and Mark V. Pauly, "Accountable Care Organizations May Have Difficulty Avoiding the Failures of Integrated Delivery Networks of the 1990s," *Health Affairs 31* (2012): 2407–2415.

Burstein, Philip L. and Jerry Cromwell, "Relative Incomes and Rates of Return for U.S. Physicians," *Journal of Health Economics 4* (1985): 63–78.

Calem, Paul S., and John A. Rizzo, "Competition and Specialization in the Hospital Industry: An Application of Hotelling's Location Model," *Southern Economic Journal 61* (1995): 1182–1198.

Cameron, Sam, "Are Greek Smokers Rational Addicts?" *Applied Economics Letters 4* (1997): 401–402.

Campbell, Maude, "Looking Down the Pharmaceutical Pipeline," *American Health & Drug Benefits* (June 2008): 37–38.

Cannon, Michael, "Large Health Savings Accounts: A Step Toward Tax Neutrality for Health Care," *Forum for Health Economics & Policy 11* (2008): Article 3.

Capps, Cory, David Dranove, and Richard C. Lindrooth, "Hospital Closure and Economic Efficiency," *Journal of Health Economics 29* (2010): 87–109.

Card, David, Carlos Dobkin, and Nicole Maestas, "Does Medicare Save Lives?" *Quarterly Journal of Economics*, *124* (2009): 597–636.

Cardon, James H., and Igal Hendel, "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *RAND Journal of Economics 23* (2001): 408–427.

Carey, Kathleen, and James F. Burgess Jr., "On Measuring the Hospital Cost/Quality Trade Off," *Health Economics 8* (1999): 509–520.

Carey, Kathleen, and Theodore Stefos, "Controlling for Quality in the Hospital Cost Function," *Health Care Management Science 14* (2011): 125–134.

Carlsen, Fredrik, and Jostein Grytten, "Consumer Satisfaction and Supplier Induced Demand," *Journal of Health Economics 19* (2000): 731–783.

Case, Meredith A., et al., "Accuracy of Smartphone Applications and Wearable Devices for Tracking Physical Activity Data," *Journal of the American Medical Association 313* (6) (2015): 625–626.

Caudill, Stephen B., John M. Ford, and David L. Kaserman, "Certificate of Need Regulation and the Diffusion of Innovations," *Journal of Applied Econometrics 10* (1995): 73–78.

Cawley, John, "An Economy of Scales: A Selective Review of Obesity's Economic Causes, Consequences, and Solutions," *Journal of Health Economics 43* (2015): 244–268.

Cawley, John, "The Impact of Obesity on Wages," *Journal of Human Resources 39* (2004): 451–474.

Cawley, John, et al., "Savings in Medical Expenditures Associated with Reductions in Body Mass Index among U.S. Adults with Obesity, by Diabetes Status," *Pharmacoeconomics 33* (2015): 707–722.

The Center for Evaluative Clinical Studies, *Dartmouth Atlas of Health Care* (Dartmouth Medical School, American Hospital Publications, Inc., 1996).

Central Intelligence Agency, *CIA World Factbook*, various years.

Chakvarty, Sujoy, et al., "Does the Profit Motive Make Jack Nimble? Ownership Form and the Evolution of the U.S. Hospital Industry," *NBER* Working Paper 11705, 2005.

Chaloupka, Frank, "The Economics of Smoking," in *Handbook of Health Economics*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam: Elsevier Science B.V., 2000).

Chaloupka, Frank J., "Rational Addictive Behavior and Cigarette Smoking," *Journal of Political Economy 99* (1991): 722–742.

Chaloupka, Frank J., and Rosalie Liccardo Pacula, "An Examination of Gender and Race Differences in Youth Smoking Responsiveness to Price and Tobacco Control Policies," *NBER* Working Paper 6541, 1998.

Chaloupka, Frank J., and Henry Saffer, "Clean Indoor Air Laws and the Demand for Cigarettes," *Contemporary Policy Issues 10* (1992): 72–83.

Chaloupka, Frank J., and Kenneth E. Warner, "The Economics of Smoking," *NBER* Working Paper 7047, 1999.

Chandra, Amitabh, David Cutler, and Zirui Song, "Who Ordered That? The Economics of Treatment in Medical Care," in *Handbook of Health Economics*, Volume 2, Mark V. Pauly, Thomas McGuire, and Pedro P. Barros (eds.) (Amsterdam: Elsevier Science B.V., 2012): 397–432.

Chandra, Amitabh, Jonathan Gruber, and Robin McKnight, "Patient Cost-Sharing and Hospitalization Offsets in the Elderly," *American Economic Review 100* (2010): 193–213.

Chang, Ji Eun, Suzanne Brundage, and Dave A. Cholski, "Convenient Ambulatory Care—Promise, Pitfalls, and Policy," *New England Journal of Medicine 373* (2015): 382–388.

Chassin Mark R., and Sally M. McCue, "A Randomized Trial of Medical Quality Assurance. Improving Physicians' Use of Pelvimetry," *Journal of the American Medical Association 256* (1986): 1012–1016.

Chatterji, Pinka, and Jeff DeSimone, "Adolescent Drinking and High School Dropout," *NBER* Working Paper 11337, 2005.

Chen, Lin Arjan de Haan, Zhang Xiulan, and Ward Warmerdam, "Addressing Vulnerability in an Emerging Economy: China's New Cooperative Medical Scheme (NCMS)," Institute of Social Studies, Erasmus University, 2011.

Chernew, Michael, and Dennis P. Scanlon, "Health Plan Report Cards and Insurance Choice," *Inquiry 35* (1998): 9–22.

Chilingerian, Jon, "Evaluating Physician Efficiency in Hospitals: A Multivariate Analysis of Best Practices," *European Journal of Operations Research 80* (1995): 548–574.

Chiswick, Barry R., "The Demand for Nursing Home Care: An Analysis of the Substitution between Institutional and Noninstitutional Care," *Journal of Human Resources 11* (1976): 295–316.

Chou, Shin-Yi, "Asymmetric Information, Ownership and Quality of Care: An Empirical Analysis of Nursing Homes," *Journal of Health Economics 21* (2002): 293–311.

Christakis, Nicholas A., and Fowler, J., "Collective Dynamics of Smoking in a Large Social Network," *New England Journal of Medicine 358* (2008): 2249–2258.

Christianson, Jon B., et al., "The HMO Industry: Evolution in Population Demographics and Market Structures," *Medical Care Review 48* (1991): 3–46.

Christensen, Eric W., and Richard J. Arnould, "The Impact of Asymmetric Information and Ownership on Nursing Home Access," *International Journal of Health Care Finances and Economics 5* (2005): 273–297.

Civan, Abdulkadir, and Bülent Köksal, "The Effect of Newer Drugs on Health Spending: Do They Really Increase the Costs?" *Health Economics 19* (2010): 581–595.

Clancy, Carolyn, and Anne K. Gauthier (eds.), "Consumer Driven Health Care: Beyond Rhetoric with Research and Experience," *Health Services Research 30* Suppl. (2004): 1049–1233.

Clark, Andrew E., and Youenn Loheac, "'It Wasn't Me, It Was Them!' Social Influence in Risky Behavior by Adolescents," *Journal of Health Economics 26* (2007): 763–784.

Clark, Robert Charles, "Does the Nonprofit Form Fit the Hospital Industry?" *Harvard Law Review 93* (1980): 1417–1489.

Clarkson, Kenneth W., "The Effects of Research and Promotion on Rates of Return," in *Competitive Strategies in the Pharmaceutical Industry*, Robert B. Helms (ed.) (Washington, DC: American Enterprise Institute, 1996) 238–268.

Claxton, Karl, et al., "Discounting and Cost-effectiveness in NICE—Stepping Back to Sort Out a Confusion," *Health Economics 15* (2006): 1–4.

Clemens, Jeffrey, and David M. Cutler, "Who Pays for Public Employee Health Costs?" *Journal of Health Economics 38* (2014) 65–76.

Cleverly, William O., "More Efficient Hospitals Are Closing," *Healthcare Financial Management 47* (1993): 82–85.

Cochrane, Archibald L., *Effectiveness and Efficiency: Random Reflections on Health Services* (London: Nuffield Provincial Hospital Trust, 1972).

Coffey, Rosanna M., "The Effect of Time Price on the Demand for Medical Care Services," *Journal of Human Resources 18* (1983): 407–424.

Cogan, John F., R. Glenn Hubbard, and Daniel P. Kessler, *Healthy, Wealthy, & Wise: Five Steps to a Better Health Care System* (Washington, DC: American Enterprise Institute and Stanford, CA: Hoover Institution, 2005).

Cohen, Joel W., and William D. Spector, "The Effect of Medicaid Reimbursement on Quality of Care in Nursing Homes," *Journal of Health Economics 15* (1996): 23–48.

Cohen, Joshua T., Peter J. Neumann, and Milton C. Weinstein, "Does Preventive Care Save Money? Health Economics and the Presidential Candidates," *New England Journal of Medicine 358* (2008): 661–663.

Coleman, James S., "Social Capital in the Creation of Human Capital," *American Journal of Sociology* Supplement *94* (1988): S95–S120.

Colla, Carrie H., et al., "Spending Differences Associated with Medicare Physician Group Practice Demonstration," *Journal of the American Medical Association 308* (2012): 1015–1023.

Collins, Robert, and Gregory A. Baker, "Of Junk Food and Junk Science," *International Food and Agribusiness Management Review 12* (2009): 111–125.

Collins, Sara R., et al., "Losing Ground: How the Loss of Adequate Health Insurance Is Burdening Working Families," Findings from the Commonwealth Fund Biennial Health Insurance Surveys, 2001–2007, The Commonwealth Fund, August 2008a.

Collins, Sara R., et al., "The 2008 Presidential Candidates' Health Reform Proposals: Choices for America," The Commonwealth Fund, October 2008b.

Comanor, William S., and Thomas A. Wilson, *Advertising and Market Power* (Cambridge, MA: Harvard, 1974).

Comanor, William S., and Thomas A. Wilson, "The Effect of Advertising on Competition: A Survey," *Journal of Economic Literature 17* (1979): 453–476.

Comanor, William S., and F.M. Scherer, "Mergers and Innovation in the Pharmaceutical Industry," *Journal of Health Economics 32* (2013): 106–113.

Comanor, William, H.E. Frech III, and Richard Miller, "Is the United States an Outlier in Health Care and Health Outcomes? A Preliminary Analysis," *International Journal of Health Care Finance and Economics 6* (2006): 3–33.

Commission for Healthcare Audit and Inspection, "State of Healthcare 2005," www.healthcarecommission.org.uk/healthcareproviders/nationalfindings/publications/publications2004-05, accessed August 2, 2008.

Commonwealth Fund Commission on a High Performance Health System, *Why Not the Best? Results from the National Scorecard on U.S. Health System Performance, 2008*, July 27, 2008. Available at www.commonwealthfund.org/publications/publications_show.htm?doc_id=692682, accessed November 2016.

Conrad, Robert F., and Robert P. Strauss, "A Multiple-Output Multiple-Input Model of the Hospital Industry in North Carolina," *Applied Economics 15* (1983): 341–352.

Contoyannis, Paul, and Andrew Jones, "Socio-Economic Status, Health and Lifestyle," *Journal of Health Economics 23* (2004): 965–995.

Conway, Karen S., and Andrea Kutinova, "Maternal Health: Does Prenatal Care Make a Difference?" *Health Economics 15* (2006): 461–488.

Cook, Philip J., and Christopher Carpenter, "Cigarette Taxes and Youth Smoking: New Evidence from National, State and Local Youth Behavior Surveys," *Journal of Health Economics 27* (2008): 287–299.

Cook, Philip J., and Michael J. Moore, "Alcohol," in *Handbook of Health Economics*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam: Elsevier Science B.V., 2000).

Cook, Philip J., and Michael J. Moore, "Drinking and Schooling," *Journal of Health Economics 12* (1993): 411–429.

Cook, Phillip J., Jan Orterman, and Frank A. Sloan, "Are Alcohol Excise Taxes Good for Us? Short and Long Term Effects on Mortality Rates," *NBER* Working Paper 11139, 2005.

Cookson, Richard, "QALYs and the Capability Approach," *Health Economics 14* (2005): 817–830.

Cooper, Philip F., and Alan C. Monheit, "Does Employment-Related Health Insurance Inhibit Job Mobility?" *Inquiry 30* (1993): 400–416.

Cooper, Philip F., Kosali I. Simon, and Jessica Vistnes, "A Closer Look at the Managed Care Backlash," *Medical Care 44* (2006): I4–I11.

Cooper, Richard A., et al., "Economic and Demographic Trends Signal an Impending Physician Shortage," *Health Affairs 21* (2002): 140–154.

Corman, Hope, and Michael Grossman, "Determinants of Neonatal Mortality Rate in the U.S.: A Reduced Form Model," *Journal of Health Economics 4* (1985): 213–236.

Corman, Hope, Theodore J. Joyce, and Michael Grossman, "Birth Outcome Production Function in the United States," *Journal of Human Resources 22* (1987): 339–360.

Cost-Effectiveness Analysis Registry (2008), https://research.tufts-nemc.org/cear/default.aspx, accessed November 2016.

Courtemanche, Charles, and Michael Plotzke, "Does Competition from Ambulatory Surgical Centers Affect Hospital Surgical Output?" *Journal of Health Economics 29* (2010): 765–773.

Coughlin, Teresa A., John Holahan, Kyle Caswell, and Megan McGrath, *Uncompensated Care for Uninsured in 2013: A Detailed Examination*, The Urban Institute, May 2014, https://kaiserfamilyfoundation.files.wordpress.com/2014/05/8596-uncompensated-care-for-the-uninsured-in-2013.pdf, accessed June 20, 2016.

Cowing, Thomas G., and Alphonse G. Holtmann, "Multiproduct Short-Run Hospital Cost Functions: Empirical Evidence and Policy Implications from Cross-Section Data," *Southern Economic Journal 49* (1983): 637–653.

Coyle, Douglas, Martin J. Buxton, and Bernie J. O'Brien, "Stratified Cost-Effectiveness Analysis: A Framework for Establishing Efficient Limited Criteria," *Health Economics 12* (2003): 421–427.

Cremieux, Pierre-Yves, and Pierre Ouellette, "Omitted Variable Bias and Hospital Costs," *Journal of Health Economics 20* (2001): 271–282.

Cremieux, Pierre-Yves, Pierre Ouellette, and Caroline Pilon, "Health Care Spending as Determinants of Health Outcomes," *Health Economics 8* (November 1999): 627–639.

Cropper, Maureen L., et al., "The Health Benefit of Air Pollution Control in New Delhi," *American Journal of Agricultural Economics 79* (1997): 1625–1629.

Cropper, Maureen, James K. Hammit, and Lisa A. Robinson, "Valuing Mortality Risk Reductions: Progress and Challenges," Discussion Paper, Washington, DC, Resources for the Future, (2011).

Culyer, Anthony J., "The Normative Economics of Health Care Finance and Provision," *Oxford Review of Economic Policy 5* (1989): 34–58.

Culyer, Anthony J., "The United Kingdom: Effective, Efficient, Equitable?" *Journal of Health Politics, Policy and Law 17* (1992): 667–688.

Culyer, Anthony J., and Adam Wagstaff, "Equity and Equality in Health and Health Care," *Journal of Health Economics 12* (1993): 431–457.

Culyer, Anthony J., Eddy van Doorslaer, and Adam Wagstaff, "Access, Utilisation and Equity: A Further Comment," *Journal of Health Economics 11* (1992a): 207–210.

Culyer, Anthony J., Eddy van Doorslaer, and Adam Wagstaff, "Utilisation as a Measure of Equity by Mooney, Hall, Donaldson and Gerard," *Journal of Health Economics 11* (1992b): 93–98.

Currie, Janet., "The Take-Up of Social Benefits," in *Poverty, the Distribution of Income, and Public Policy*, Alan Auerbach, David Card, and John M. Quigley (eds.) (New York: Russell Sage, 2006): 80–148.

Currie, Janet, and Nancy Cole, "Welfare and Child Health: The Link between AFDC Participation and Birth Weight," *American Economic Review 83* (1993): 971–985.

Currie, Janet, and Jonathan Gruber, "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women," *Journal of Political Economy 104* (1996): 1263–1296.

Currie, Janet, and Patricia B. Reagan, "Distance to Hospital and Children's Use of Preventive Care: Is Being Closer Better, and for Whom?" *Economic Inquiry 41* (2005): 378–391.

Currie, Janet, et al., "The Effect of Fast Food Restaurants on Obesity and Weight Gain," *American Economic Review: Economic Policy 2* (2010): 32–63.

Custer, W. S., and Richard J. Willke, "Teaching Hospital Costs: The Effect of Medical Staff Characteristics," *Health Services Research 25* (1991): 831–857.

Cutler, David M., "The Lifetime Costs and Benefits of Medical Technology," *Journal of Health Economics 26* (2007): 1081–1100.

Cutler, David M., *Your Money or Your Life: Strong Medicine for America's Health Care System* (Oxford: Oxford University Press, 2004).

Cutler, David M., and Jonathan Gruber, "Does Public Insurance Crowd Out Private Insurance," *Quarterly Journal of Economics 111* (1996): 391–430.

Cutler, David M., and Jonathan Gruber, "Health Policy in the Clinton Era: Once Bitten, Twice Shy," in *American Economic Policy in the 1990s*, Peter R. Orszag and Jeffrey A. Frankel (eds.) (Cambridge, MA: MIT Press, 2002).

Cutler, David M., and Jill R. Horwitz, "Converting Hospitals from Not-for-Profit to For-Profit Status: Why and What Effects?" *NBER* Working Paper 6672, 1998.

Cutler, David M., and Adriana Lleras-Muney, "Education and Health Evaluating Theories and Evidence," *NBER* Working Paper 12352, 2006.

Cutler, David M., and Dan P. Ly, "The (Paper)Work of Medicine: Understanding International Medical Costs," *Journal of Economic Perspectives 25* (2011): 3–25.

Cutler, David M., and Ellen Meara, "The Medical Costs of the Young and Old: A Forty-Year Perspective," *NBER* Working Paper 6114, 1997.

Cutler, David M., and Grant Miller, "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States," *Demography 42* (2005): 1–22.

Cutler, David M., and Sarah J. Reber, "The Trade-Off between Competition and Adverse Selection," *Quarterly Journal of Economics 113* (1998): 433–466.

Cutler, David M., Edward L. Glaeser, and Jesse M. Shapiro, "Why Have Americans Become More Obese?" *Journal of Economic Perspectives 17* (2003): 93–118.

Cutler, David M., Mark McClellan, and Joseph P. Newhouse, "How Does Managed Care Do It?" *RAND Journal of Economics 31* (2000): 526–548.

Cutler, David M., Mark McClellan, Joseph P. Newhouse, and Dahlia Remler, "Pricing Heart Attack Treatments," in David M. Cutler and Ernst R. Berndt (eds.), *Medical Care Output and Productivity* (Chicago: University of Chicago Press, 2001).

Dafny, Leemore, Mark Duggan, and Subramanian Ramanarayanan, "Paying a Premium on Your Premium? Consolidation in the U.S. Health Insurance Industry," *American Economic Review 102* (2012): 1161–1185.

Dafny, Leemore, Kate Ho, and Robin S. Lee, "The Price Effects of Cross-Market Hospital Mergers," National Bureau of Economic Research Working Paper No. 22106, March 2016.

Dalton, Kathleen, and Hilda A. Howard, "Market Entry and Exit in Long-Term Care: 1985–2000," *Health Care Financing Review 24* (2002): 17–32.

Daniels, Norman, "Equity of Access to Health Care: Some Conceptual and Ethical Issues," *Milbank Memorial Fund Quarterly/Health and Society 60* (1982): 51–81.

Daniels, Norman, "Health Care Needs and Distributive Justice," *Philosophy and Public Affairs 10* (1981): 146–179.

Daniels, Norman, *Just Health Care* (Cambridge: Cambridge University Press, 1985).

Danzon, Patricia M., Y. Richard Wang, and Liang Wang, "The Impact of Price Regulation on the Launch Delay of New Drugs—Evidence from Twenty-Five Major Markets in the 1990s," *Health Economics 14* (2005): 269–292.

David, Guy, "The Convergence between For-Profit and Nonprofit Hospitals in The United States," *International Journal of Health Care Finance and Economics 9* (2009): 403–428.

David, Guy, and Scott Harrington, "Population Density and Racial Differences in the Performance of Emergency Medical Services," *Journal of Health Economics 29* (2010): 603–615.

Davis, Adam J., Karl R. Geisler, and Mark W. Nichols, "The Price Elasticity of Marijuana Demand: Evidence from Crowd-Sourced Transactions Data," *Empirical Economics 50* (2015): 1–22.

Davis, Mark A., "On Nursing Home Quality: A Review and Analysis," *Medical Care Review 48* (1991): 129–166.

Deaton, Angus, and Christina Paxson, "Mortality, Income, and Income Inequality over Time in Britain and the United States," *NBER* Working Paper 8534, 2001.

Deb, Partha M., and Ann M. Holmes, "Substitution of Physicians and Other Providers in Outpatient Mental Health Care," *Health Economics 7* (1998): 347–361.

Deb, Partha M., et al., "The Effect of Managed Care on Use of Health Care Services: Results from Two Contemporaneous Household Surveys," *Health Economics 15* (2006): 743–760.

Decker, Frederic H., "Nursing Home Performance in Resident Care in the United States: Is It Only a Matter of For-Profit versus Not-For-Profit?" *Health Economics, Policy and Law 3* (2008): 115–140.

Defelice, Lisa C., and W. David Bradford, "Relative Inefficiencies in Production between Solo and Group Practice Physicians," *Health Economics 6* (1997): 455–465.

Deily, Mary, and Niccie McKay, "Cost Inefficiency and Mortality Rates in Florida Hospitals," *Health Economics 115* (2006): 419–431.

Dekker, Thijs, et al., "The Effect of Risk Context on the Value of a Statistical Life: A Bayesian Meta-model," *Environmental and Resource Economics 49* (2011): 597–624.

Deloitte Centre for Health Solutions, *Measuring the Return on Pharmaceutical Innovation 2014: Turning a Corner?*, Deloitte LLP (United Kingdom), 2014.

DeSalvo, Karen B., et al. "Public Health 3.0: Time for an Upgrade," *American Journal of Public Health 106* (April 2016): 621–622.

Deutsch, Larry, "Pharmaceuticals: The Critical Role of Innovation," in *Industry Studies*, 2nd ed., Larry L. Deutsch (ed.) (Armonk, NY: M.E. Sharpe, 1998).

D'Hombres, Béatrice, et al., "Does Social Capital Determine Health? Evidence from Eight Transition Countries," *Health Economics 19* (2010): 56–74.

DiCicca, Phillip, et al., "Youth Smoking, Cigarette Prices and Anti-Smoking Sentiment," *NBER* Working Paper 12458, 2006.

Diehr, Paula, et al., "What Is Too Much Variation: The Null Hypothesis in Small-Area Analysis," *Health Services Research 24* (1990): 741–771.

DiFranza, Joseph R. et al., "RJR Nabisco's Cartoon Camel Promotes Camel Cigarettes to Children," *Journal of the American Medical Association 266* (December 11, 1991): 3149–3153.

Dill, Michael J., et al., "Survey Shows Consumers Open to a Greater Role for Physician Assistants and Nurse Practitioners," *Health Affairs 32* (2013): 1135–1142.

Dills, Angela K., Mirelle Jacobson, and Jeffrey A. Miron, "The Effect of Alcohol Prohibition on Alcohol Consumption: Evidence from Drunkenness Arrests," *Economics Letters 86* (2005): 279–284.

DiMasi, Joseph A., "New Drug Development in the United States from 1963 to 1999," *Clinical Pharmacology & Therapeutics 16* (2001): 286–296.

DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen, "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs," Tufts Center for the Study of Drug Development Briefing, "Cost of Developing a New Drug," November 18, 2014.

DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski, "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics 22* (2003): 151–185.

DiMasi, Joseph A., Ronald W. Hansen, and Henry G. Grabowski, "Reply: Extraordinary Claims Require Extraordinary Evidence," *Journal of Health Economics 24* (2005): 1034–1044.

DiMasi, Joseph A. et al., "Cost of Innovation in the Pharmaceutical Industry," *Journal of Health Economics 10* (1991): 107–142.

Dixon, Anna, Jessica Greene, and Judith Hibbard, "Do Consumer-Directed Health Plans Drive Change in Enrollees' Health Care Behavior?" *Health Affairs 27* (2008): 1120–1131.

Dobkin, Carlos, and Steven L. Puller, "The Effects of Government Transfers on Monthly Cycles in Drug Abuse, Hospitalization and Mortality," *Journal of Public Economics 91* (2007): 2137–2157.

Dolan, Paul, et al., "QALY Maximization and People's Preferences: A Methodological Review of the Literature," *Health Economics 14* (2005): 197–208.

Donabedian, Avedis, *The Definition of Quality and Approaches to Its Assessment* (Ann Arbor, MI: Health Administration Press, 1980).

Donabedian, Avedis, "Evaluating the Quality of Medical Care," *Milbank Memorial Fund Quarterly/Health and Society 44* (1966): 166–203.

Donohue, Julie, "A History of Drug Advertising: The Evolving Role of Consumers and Consumer Protection," *The Milbank Quarterly 84* (2006): 659–699.

Dor, Avi, "Non-Minimum Cost Functions and the Stochastic Frontier: On Applications to Health Care Providers," *Journal of Health Economics 13* (1994): 329–334.

Doyle, Joseph, "Health Insurance, Treatment and Outcomes: Using Auto Accidents as Health Shocks," *The Review of Economics and Statistics 87* (2005): 256–270.

Doyle, Joseph J., Jr., Steven M. Ewer, and Todd H. Wagner, "Returns to Physician Human Capital: Evidence from Patients Randomized to Physician Teams," *Journal of Health Economics 29* (2010): 866–882.

Dranove, David, "Demand Inducement and the Physician/Patient Relationship," *Economic Inquiry 26* (1988): 281–298.

Dranove, David, *The Economic Evolution of American Health Care* (Princeton, NJ: Princeton University Press, 2000).

Dranove, David, "A Problem with Consumer Surplus Measures of the Cost of Practice Variations," *Journal of Health Economics 14* (1995): 243–251.

Dranove, David, and Kenneth Cone, "Do State Rate Setting Regulations Really Lower Hospital Expenses?" *Journal of Health Economics 4* (1985): 159–165.

Dranove, David, and Richard Lindrooth, "Hospital Consolidation and Costs: Another Look at the Evidence," *Journal of Health Economics 22* (2003): 983–997.

Dranove, David, and David Meltzer, "Do Important Drugs Reach the Market Sooner?" *RAND Journal of Economics 25* (1994): 402–423.

Dranove, David, and Paul Wehner, "Physician Induced Demand for Childbirths," *Journal of Health Economics 13* (1994): 61–73.

Dranove, David, and William D. White, "Agency and the Organization of Health Care Delivery," *Inquiry 24* (1987): 405–415.

Dranove, David, Carol J. Simon, and William D. White, "Determinants of Managed Care Penetration," *Journal of Health Economics 17* (1998): 729–745.

Dranove, David, Carol J. Simon, and William D. White, "Is Managed Care Leading to Consolidation in Health-care Markets?" *Health Services Research 37* (2002): 573–594.

Dranove, David et al., "Is the Impact of Managed Care on Hospital Prices Decreasing?" *Journal of Health Economics 27* (2008): 362–376.

Drummond, Michael F. et al., *Methods for the Economic Evaluation of Health Care Programmes*, 2nd ed. (New York: Oxford University Press, 1997).

Duarte, Fabian, "Price Elasticity of Expenditure across Health Care Services," *Journal of Health Economics 31* (2012): 824–841.

Dubos, Rene, *The Mirage of Health* (London: George Allen and Unwin, Ltd., 1960).

Duclos, Jean-Yves, and Damien Échevin, "Health and Income: A Robust Comparison of Canada and the US," *Journal of Health Economics 30* (2011): 293–302.

Duffy, Thomas P., "The Flexner Report—100 Years Later," *Yale Journal of Biology and Medicine 84* (2011): 269–276.

Dyck, Frank J., et al., "Effect of Surveillance on the Number of Hysterectomies in the Province of Saskatchewan," *New England Journal of Medicine 296* (1977): 1326–1328.

Eakin, B. Kelly, and Thomas J. Kniesner, "Estimating a Non-Minimum Cost Function for Hospitals," *Southern Economic Journal 54* (1988): 583–597.

Easley, David, and Maureen O'Hara, "The Economic Role of the Nonprofit Firm," *Bell Journal of Economics 14* (1983): 531–538.

Eddy, David M., "Clinical Decision Making: From Theory to Practice," *Journal of the American Medical Association 263* (January 19, 1990): 441–443.

Efron, Bradley and Robert Tibshirani, *An Introduction to the Bootstrap* (London: Chapman and Hall, 1993).

Eggleston, Karen, "Health Care for 1.3 Billion: An Overview of China's Health System," Stanford University, January 2012, http://aparc.fsi.stanford.edu/publications/health_care_for_13_billion_an_overview_of_chinas_health_system, accessed June 3, 2016.

Eggleston, Karen, et al., "Health Service Delivery in China: A Literature Review", *Health Economics 17* (2008a): 149–165.

Eggleston, Karen, et al., "Hospital Ownership and Quality of Care: What Explains the Different Results in the Literature?" *Health Economics 17* (2008b): 1345–1362.

Ellis, Randall P., "Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins," *Journal of Health Economics 17* (1998): 537–555.

Ellis, Randall P., "Hospital Cost Function Estimation When Firms May Not Try to Minimize Total Costs," mimeo (Boston: Boston University, 1993).

Ellis, Randall P., and Thomas G. McGuire, "Provider Behavior under Prospective Reimbursement," *Journal of Health Economics 5* (1986): 129–151.

Emanuel, Ezekiel J., and Victor R. Fuchs, "Who Really Pays for Health Care? The Myth of 'Shared Responsibility,'" *Journal of the American Medical Association 299* (2008): 1057–1059.

Epstein, Andrew J., and Sean Nicholson, "The Formation and Evolution of Physician Treatment Styles: An Application to Cesarean Sections," *Journal of Health Economics 28* (2009): 1126–1140.

Epstein, Arnold S., et al., "Analysis of Early Accountable Care Organizations Defined Patient, Structural, Cost, and Quality-of-Care Characteristics," *Health Affairs 33* (2014): 95–102.

Escarce, José J., "Externalities in Hospitals and Physicians Adoption of a New Surgical Technology: An Exploratory Analysis," *Journal of Health Economics 15* (1996): 715–734.

Escarce, José J., "Would Eliminating Differences in Physician Practice Style Reduce Geographic Variations in Cataract Surgery Rates?" *Medical Care 12* (1993): 1106–1118.

Escarce, José J., and Mark V. Pauly, "Physician Opportunity Costs in Physician Practice Cost Functions," *Journal of Health Economics 17* (1998): 129–151.

Escarce, José J., et al., "Diffusion of Laparoscopic Cholecystectomy among General Surgeons in the United States," *Medical Care 33* (1995): 256–271.

Evans, Robert G., "Behavioral Cost Functions for Hospitals," *Canadian Journal of Economics 4* (1971): 198–215.

Evans, Robert G., "Supplier-Induced Demand: Some Empirical Evidence and Implications," in *The Economics of Health and Medical Care*, Mark Perlman (ed.) (London: Macmillan, 1974).

Evans, William N., and Diana S. Lien, "The Benefits of Prenatal Care: Evidence from the PAT Bus Strike," *Journal of Econometrics 125* (March 2005): 207–239.

Evans, William N., and Jeanne S. Ringel, "Can Higher Cigarette Taxes Improve Birth Outcome?" *Journal of Public Economics 72* (1999): 135–154.

Fang, Hai, and John A. Rizzo, "Has the Influence of Managed Care Waned? Evidence from the Market for Physician Services," *International Journal of Health Care Finance and Economics 10* (2010): 85–103.

Fang, Hai, Hong Liu, and John A. Rizzo, "Has the Use of Physician Gatekeepers Declined among HMOs? Evidence from the United States," *International Journal of Health Care Finance and Economics 9* (2009): 183–195.

Farahani, Mansour, S. V. Subrmanian, and D. Canning, "Effects of State-Level Public Spending on Health on the Mortality Probability in India," *Health Economics 19* (2010): 1361–1376.

Fare, Rolf, and C. A. Knox Lovell, "Measuring the Technical Efficiency of Production," *Journal of Economic Theory 19* (1978): 150–162.

Farmer, Paul, and Edward Nardell, "Editorial: Nihilism and Pragmatism in Tuberculosis Control," *American Journal of Public Health 88* (1998): 1014–1015.

Farsi, Mehdi, "Changes in Hospital Quality after Conversion in Ownership Status," *International Journal of Health Care Finance and Economics 4* (2004): 211–230.

Farsi, Mehdi, "The Temporal Variation of Cost-Efficiency in Switzerland's Hospitals: An Application of Mixed Models," *Journal of Productivity Analysis*, *30* (2008): 155–168.

Farsi, Mehdi, and Massimo Philippini, "Effects of Ownership, Subsidization and Teaching: Activities on Hospital Costs in Switzerland," *Health Economics 17* (2008): 335–350.

Feldman, Roger, and Bryan Dowd, "Is There a Competitive Market for Hospital Services?" *Journal of Health Economics 5* (1986): 277–292.

Feldman, Roger, and Bryan Dowd, "A New Estimate of the Welfare Loss of Excess Health Insurance," *American Economic Review 81* (1991): 297–301.

Feldman, Roger, Stephen T. Parente, and Jon B. Christianson, "Consumer-Directed Health Plans: Evidence on Spending and Utilization," *Inquiry 44* (2007): 26–39.

Feldman, Roger, et al., "Effects of HMOs on the Creation of Competitive Markets for Hospital Services," *Journal of Health Economics 9* (1990): 207–222.

Feldstein, Martin S., *Economic Analysis for Health Service Efficiency* (Amsterdam: North-Holland, 1967).

Feldstein, Martin S., "Hospital Cost Inflation: A Study of Nonprofit Price Dynamics," *American Economic Review 61* (1971): 853–872.

Feldstein, Martin S., "The Rising Price of Physicians' Services," *Review of Economics and Statistics 52* (1970): 121–133.

Feldstein, Martin S., "The Welfare Loss of Excess Health Insurance," *Journal of Political Economy 81* (1973): 251–280.

Feldstein, Paul J., "The Emergence of Market Competition in the U.S. Health Care System. Its Causes, Likely Structure, and Implications," *Health Policy 6* (1986): 1–20.

Finkelstein, Amy, "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare," *Quarterly Journal of Economics 122* (2007): 1–37.

Finkelstein, Amy, and Robin McKnight, "What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending," *Journal of Public Economics 92* (2008): 1644–1669.

Finkelstein, Amy, et al., "The Oregon Health Insurance Experiment: Evidence from the First Year," *Quarterly Journal of Economics 127* (2012): 1057–1106.

Fiorillo, Damiano, and Fabio Sabatini, "Quality and Quantity: The Role of Social Interactions on Individual Health," *Social Science & Medicine 73* (2011): 144–1652.

Fisher, Elliott S., et al., "Creating Accountable Care Organizations: The Extended Hospital Medical Staff," *Health Affairs 26* (2006): w44–w57.

Fisher, Elliott S., et al., "A Framework for Evaluating the Formation, Implementation, and Performance of Accountable Care Organizations," *Health Affairs 31* (2012): 2368–2378.

Fletcher, Suzanne W., and Graham A. Colditz, "Failure of Estrogen Plus Progestin Therapy for Prevention," *Journal of the American Medical Association 288* (2002): 366–368.

Flexner, Abraham, *Medical Education in the United States and Canada, Bulletin No. 4* (New York: Carnegie Foundation for the Advancement of Teaching, 1910).

Fogel, Robert W., *The Escape from Hunger and Premature Death, 1700–2100: Europe, America and the Third World* (Cambridge: Cambridge University Press, 2004).

Fogel, Robert W., *The Fourth Great Awakening & the Future of Egalitarianism* (Chicago: University of Chicago Press, 2000).

Fogel, Robert, and Stanley Engerman, *Explaining Long Term Trends in Health and Longevity* (Cambridge: Cambridge University Press, 2012).

Folland, Sherman T., "A Critique of Pure Need: An Analysis of Norman Daniels's Concept of Health Care Need," *International Journal of Social Economics 17* (1990): 36–45.

Folland, Sherman, "Does Community Social Capital Contribute to Population Health?" *Social Science & Medicine 64* (2007): 2342–2354.

Folland, Sherman, "An Economic Model of Social Capital and Health," *Health Economics, Policy and Law 3* (2008): 333–348.

Folland, Sherman T., "Health Care Need, Economics, and Social Justice," *International Journal of Social Economics 13* (1986): 98–116.

Folland, Sherman T., "The Quality of Mercy: Social Health Insurance in the Charitable Liberal State," *International Journal of Health Care Finance and Economics 5* (2005): 23–46.

Folland, Sherman, "What Is Social Capital and How Does It Work to Improve Health?" in *The Economics of Social Capital and Health*, S. Folland and L. Rocco, World Scientific (Singapore: World Scientific, 2014).

Folland, Sherman, "Value of Life and Behavior toward Health Risks: An Interpretation of Social Capital," *Health Economics 15* (2006): 159–171.

Folland, Sherman T., and Richard A. Hofler, "How Reliable Are Hospital Efficiency Estimates? Exploiting the Dual to Homothetic Production," *Health Economics 10* (2001): 683–698.

Folland, Sherman, and E. Iversen, "How Does Social Capital Arise in Populations," in *The Economics of Social and Health*, S. Folland and L. Rocco (eds.), 45–66 (Singapore: World Scientific, 2014).

Folland, Sherman, and Lorenzo Rocco (eds.), *The Economics of Social Capital and Health* (New York: World Scientific, 2014).

Folland, Sherman T., and Miron Stano, "Small Area Variations: A Critical Review of Propositions, Methods, and Evidence," *Medical Care Review 47* (1990): 419–465.

Folland, Sherman T., and Miron Stano, "Sources of Small Area Variations in the Use of Medical Care," *Journal of Health Economics 8* (1989): 85–107.

Folland, Sherman, M. Kamrul Islam, and Oddvar Kaarboe, "The Social Capital and Health Hypothesis: An Integrating Theory and New Empirics," Oakland University, 2011.

Folland, Sherman, O. Kaarboe, and K. Islam, "How Do We Invest in Social Capital? An Exploration of an Economic Model of Social Capital and Health," in *The Economics of Social Capital and Health*, S. Folland and L. Rocco (eds.), 17–26 (Singapore: World Scientific, 2014).

Forder, Julien, "Long-Term Care and Hospital Utilisation by Older People: An Analysis of Substitution Rates," *Health Economics 18* (2009): 1332–1338.

Fournier, Gary M., and Jean M. Mitchell, "Hospital Costs and Competition for Services: A Multi-Product Analysis," *Review of Economics and Statistics 74* (1992): 627–634.

Fox, William F., and Richard A. Hofler, "Using Homothetic Composed Error Frontiers to Measure Water Utility Efficiency," *Southern Economic Journal 53* (1986): 461–477.

Frank, Richard G., "Editorial: New Estimates of Drug Development Costs," *Journal of Health Economics 22* (2003): 325–330.

Frank, Richard G., and David S. Salkever, "Generic Entry and the Pricing of Pharmaceuticals," *Journal of Economics and Management Strategy 6* (1997): 75–90.

Frank, Richard G., and David S. Salkever, "Nonprofit Organizations in the Health Sector," *Journal of Economic Perspectives 8* (1994): 129–144.

Frean, Molly, Jonathan Gruber, and Benjamin D. Sommers "Premium Subsidies, the Mandate, and Medicaid Expansion: Coverage Effects of the Affordable Care Act," *NBER* Working Paper 22213 www.nber.org/papers/w22213, June 2016.

Frech, H.E., III, *Competition & Monopoly in Medical Care* (Washington, DC: AEI Press, 1996).

Frech, H.E., III, *The United States, Advances in Health Economics and Health Services Research Supplement 1: Comparative Health Systems* (Greenwich, CT: JAI Press, 1990): 43–75.

Frech, H.E., III and Paul B. Ginsburg, "Competition among Health Insurers, Revisited," *Journal of Health Politics, Policy and Law 13* (1988): 279–291.

Frech, H.E., III and Richard D. Miller Jr., *The Productivity of Health Care and Pharmaceuticals* (Washington, DC: American Enterprise Institute, 1999).

Frederick, Shane, George Loewenstein, and Ted O'Donoghue, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature 40* (2002): 351–401.

Frenk, Julio, "Editorial: Medical Care and Health Improvement: The Critical Link," *Annals of Internal Medicine 129* (1998): 419–420.

Friedman, Bernard, and Claudia Steiner, "Does Managed Care Affect the Supply and Use of ICU Services?" *Inquiry 36* (1999): 68–77.

Friedman, Milton, and Simon Kuznets, *Income from Independent Professional Practice* (New York: National Bureau of Economic Research, 1945).

Fuchs, Victor R., "The Clinton Plan: A Researcher Examines Reform," *Health Affairs 13* (1994): 102–114.

Fuchs, Victor R., "Economics, Health, and Post Industrial Society," *Milbank Memorial Fund Quarterly/Health and Society 57* (1979): 153–182.

Fuchs, Victor R., "The Future of Health Economics," *Journal of Health Economics 19* (2000): 141–157.

Fuchs, Victor R., "Health Care for the Elderly: How Much? Who Will Pay for It?" *Health Affairs 18* (1999): 11–21.

Fuchs, Victor R., "Medicare Reform: The Larger Picture," *Journal of Economic Perspectives 14* (2000): 57–70.

Fuchs, Victor R., "The Supply of Surgeons and the Demand for Operations," *Journal of Human Resources 13* (Supplement 1978): 35–56.

Fuchs, Victor R., "Time Preference and Health: An Exploratory Study," in *Economic Aspects of Health*, Victor R. Fuchs (ed.) (Chicago: The University of Chicago Press, 1982): 93–120.

Fuchs, Victor R., *Who Shall Live? Health, Economics, and Social Choice, Expanded Edition* (Singapore: World Scientific Publishing Company, 1975).

Fuchs, Victor R., and James S. Hahn, "How Does Canada Do It? A Comparison of Expenditures for Physicians' Services in the United States and Canada," *New England Journal of Medicine 323* (1990): 884–890.

Gagne, Robert, and Pierre Thomas Leger, "Determinants of Physicians' Decisions to Specialize," *Health Economics 14* (2005): 721–735.

Gallett, Craig A., "Advertising and Restrictions in the Cigarette Industry: Evidence of State-by-State Variation," *Contemporary Economic Policy 21* (2003): 338–348.

Garber, Alan, "Advances in Cost-Effectiveness Analysis of Health Interventions," in *Handbook of Health Economics, Volume 1*, A.J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam: Elsevier Science B.V., 2000).

Garber, Alan M., and Charles E. Phelps, "Economic Foundations of Cost-Effectiveness Analysis," *Journal of Health Economics 16* (1997): 1–31.

García-Frapolli, et al., "Beyond Fuelwood Savings: Valuing the Economic Benefits of Introducing Improved Biomass Cookstoves in the Purépecha Region of Mexico," *Ecological Economics 69* (2010): 2508–2605.

Gaumer, Gary L., "Regulating Health Professionals: A Review of the Empirical Literature," *Milbank Memorial Fund Quarterly/Health and Society 62* (1984): 380–416.

Gaynor, Martin, "What Do We Know about Competition and Quality in Health Care Markets?" *NBER* Working Paper 12301, 2006.

Gaynor, Martin, and Gerard F. Andersen, "Uncertain Demand, the Structure of Hospital Costs, and the Cost of Empty Beds," *NBER* Working Paper 4460, 1995.

Gaynor, Martin, and Mark V. Pauly, "Compensation and Productive Efficiency in Partnerships: Evidence from Medical Group Practice," *Journal of Political Economy 98* (1990): 544–573.

Gaynor, Martin, and Solomon Polachek, "Measuring Information in the Market: An Application to Physician Services," *Southern Economic Journal 60* (1994): 815–831.

Gaynor, Martin, and Robert J. Town, "Competition in Health Care Markets," in *Handbook of Health Economics*, Volume 2, Mark V. Pauly, Thomas McGuire, and Pedro P. Barros (eds.), 499–638 (Amsterdam: Elsevier Science B.V., 2012).

Gaynor, Martin, and Robert Town, "The Impact of Hospital Consolidation—Update," The Synthesis Project: Robert Wood Johnson Foundation, June 2012.

Gaynor, Martin, and William B. Vogt, "Antitrust and Competition in Health Care Markets," in *Handbook of Health Economics*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam: Elsevier Science B.V., 2000).

Gaynor, Martin, Deborah Haas-Wilson, and William B. Vogt, "Are Invisible Hands Good Hands? Moral Hazard, Competition, and the Second Best in Health Care Markets," *Journal of Political Economy 108* (2000): 992–1005.

Gerdtham, Ulf-G., and M. Johannesson, "New Estimates of the Demand for Health: Results Based on a Categorical Health Measure and Swedish Micro Data," *Social Science and Medicine 49* (1999): 1325–1332.

Gerdtham, Ulf-G., and Bengt Jonsson, "International Comparisons of Health Expenditure," in *Handbook of Health Economics, Volume 1A*, Anthony J. Culyer and Joseph P. Newhouse (eds.), 11–53 (Amsterdam: Elsevier Science, B.V., 2000).

Gerdtham, Ulf-G., et al., "An Econometric Analysis of Health Care Expenditure: A Cross-Section Study of the OECD Countries," *Journal of Health Economics 11* (1992): 63–84.

Gerking, Shelby, and Raman Khaddaria, "Perceptions of Health Risks and Smoking Decisions of Young People," *Health Economics 21* (2012): 865–877.

Gertler, Paul J., "Subsidies, Quality, and the Regulation of Nursing Homes," *Journal of Public Economics 38* (1989): 33–52.

Getzen, Thomas E., "Health Care Is an Individual Necessity and a National Luxury," *Journal of Health Economics 19* (2000): 259–270.

Getzen, Thomas E., and Jean-Pierre Poullier, "International Health Spending Forecasts: Concepts and Evaluation," *Social Science and Medicine 34* (1992): 1057–1068.

Giaccotto, Carmelo, Rexford Santerre, and John A. Vernon, "Drug Prices and Research and Development Investment Behavior in the Pharmaceutical Industry," *Journal of Law and Economics 48* (2005): 195–214.

Gibson, Teresa R., Ronald J. Ozminkowski, and Ron Z. Goetzel, "The Effects of Prescription Drug Cost Sharing: A Review of the Evidence," *American Journal of Managed Care 11* (2005): 730–740.

Gillum B.S., E.J. Graves, and E. Wood, *National Hospital Discharge Survey: Annual Summary, 1995*, National Center for Health Statistics, Vital Health Statistics *13* (1998).

Ginsburg, Paul B., and Joy M. Grossman, "When The Price Isn't Right: How Inadvertent Payment Incentives Drive Medical Care," *Health Affairs (2005): W5-376–W5-384,* http://content.healthaffairs.org/content/early/2005/08/09/hlthaff.w5.376.citation, accessed November 2016.

Glaeser, Edward L., David Laibson, and Bruce Sacerdote, "An Economic Approach to Social Capital," *Economic Journal 112* (2002): F437–458.

Glick, Peter, and David E. Sahn, "Health and Productivity in a Heterogeneous Urban Labor Market," *Applied Economics 30* (1998): 203–216.

Glied, Sherry, "Managed Care," in *Handbook of Health Economics, Volume 1A*, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam: Elsevier Science, B.V., 2000).

Goddeeris, John H., and Burton A. Weisbrod, "Conversion from Nonprofit to For-Profit Legal Status: Why Does It Happen and Should Anyone Care?" *Journal of Policy Analysis and Management 17* (1998): 215–233.

Gold, Marthe R., et al. (eds.), *Cost Effectiveness in Health and Medicine* (New York: Oxford University Press, 1996).

Goldberg, Lawrence G., and Warren Greenberg, "The Effect of Physician-Controlled Health Insurance," *Journal of Health Politics, Policy and Law 2* (1977): 48–78.

Goldberg, Theodore, et al., "Evaluation of Economic Effects of Drug Product Selection Legislation," *Medical Care 17* (1979): 411–419.

Goldman, Dana P., et al., "Pharmacy Benefits and the Use of Drugs by the Chronically Ill," *Journal of the American Medical Association 291* (2004): 2344–2350.

Goldman, Dana P., et al., "Redistributional Consequences of Community Rating," *Health Services Research 32* (1997): 71–86.

Golec, Joseph, and John A. Vernon, "New Estimates of Pharmaceutical Research and Development Spending by US-based Firms from 1984 to 2003," *Managerial and Decision Economics 28* (2007): 481–483.

Gonçalves, Ricardo, and Pedro P. Barros, "Economies of Scale and Scope in the Provision of Diagnostic Techniques and Therapeutic Services in Portuguese Hospitals," *Applied Economics 45* (2013): 415–433.

Goodman, Allen C., and Miron Stano, "HMOs and Health Externalities: A Local Public Good Perspective," *Public Finance Review 28* (2000): 247–269.

Goodman, Allen C., Miron Stano, and John M. Tilford, "Applications and Extensions of the Grossman Health Care Model," *Southern Economic Journal 65* (1999): 791–806.

Goodman, Clifford S., "Healthcare Technology Assessment: Methods, Framework, and Role in Policy Making," *American Journal of Managed Care 4* (1998: Special Issue): SP200–SP213.

Gottlieb, David, and Olivia S. Mitchell, "Narrow Framing and Long-Term Insurance," mimeo, June 2015.

Grabowski, David C., "Medicare and Medicaid: Conflicting Incentives for Long-Term Care," *The Milbank Quarterly 85* (2007): 579–610.

Grabowski, David C., and Richard Hirth, "Competitive Spillovers across Nonprofit and For-Profit Homes," *Journal of Health Economics 22* (2003): 1–22.

Grabowski, Henry G., and John M. Vernon, "Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act," *Journal of Law and Economics 35* (1992): 331–350.

Grabowski, Henry G., and John M. Vernon, "Prospects for Returns to Pharmaceutical R&D under Health Care Reform," in *Competitive Strategies in the Pharmaceutical Industry*, Robert B. Helms (ed.), 194–207 (Washington, DC: American Enterprise Institute, 1996).

Grabowski, Henry G., and John M. Vernon, "Returns to R&D on New Drug Innovations in the 1980s," *Journal of Health Economics 13* (1994): 383–406.

Graddy, Elizabeth, "Interest Groups or the Public Interest—Why Do We Regulate Health Occupations?" *Journal of Health Politics, Policy and Law 16* (1991): 25–49.

Grannemann, Thomas W., Randall S. Brown, and Mark V. Pauly, "Estimating Hospital Costs: A Multiple-Output Analysis," *Journal of Health Economics 5* (1986): 107–127.

Grant, Darren, "Physician Financial Incentives and Cesarean Delivery: New Conclusions from the Healthcare Cost and Utilization Project," *Journal of Health Economics 28* (2009): 244–250.

Green, Linda V., Segei Savin, and Yina Lu, "Primary Care Physician Shortages Could be Eliminated Through Use of Teams," *Health Affairs 32* (2013): 11–19.

Greene, Jessica, et al., "Comprehension and Choice of a Consumer-Directed Health Plan: An Experimental Study," *American Journal of Managed Care 14* (2008a): 369–376.

Greene, Jessica, et al., "The Impact of Consumer-Directed Health Plans on Prescription Drug Use," *Health Affairs 27* (2008b): 1111–1119.

Grosse, Scott D., "Assessing Cost Effectiveness in Health Care: The History of the $50,000 per QALY Threshold," *Expert Review in Pharmacoeconomics and Outcomes Research 8* (2008), 165–178.

Grossman, Michael, *The Demand for Health: A Theoretical and Empirical Investigation* (New York: Columbia [for the National Bureau of Economic Research], 1972a).

Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy 80* (1972b): 223–255.

Grossman, Michael, and Theodore J. Joyce, "Unobservables, Pregnancy Resolutions, and Birth Weight Production Functions in New York City," *Journal of Political Economy 98* (1990): 983–1007.

Grossman, Michael, Frank J. Chaloupka, and Ismail Sirtalan, "An Empirical Analysis of Alcohol Addiction: Results from the Monitoring the Future Panels," *Economic Inquiry 36* (1998): 39–48.

Gruber, Jonathan, "Covering the Uninsured in the United States," *Journal of Economic Literature 46* (2008): 571–606.

Gruber, Jonathan, "Health Care Reform is a Three-Legged Stool," Center for American Progress (2010), www.americanprogress.org/issues/healthcare/report/2010/08/05/8226/health-care-reform-is-a-three-legged-stool/, accessed August 23, 2016.

Gruber, Jonathan, "The Incidence of Mandated Maternity Benefits," *American Economic Review 84* (1994): 622–641.

Gruber, Jonathan, "*Medicaid," in Means-Tested Transfers in the U.S.*, Robert A. Moffit (ed.) (Chicago: University of Chicago Press, 2002).

Gruber, Jonathan, and Daniel M. Hungerman, "Faith-Based Charity and Crowd-Out during the Great Depression," *Journal of Public Economics 91* (2007): 1043–1069.

Gruber, Jonathan, and Botond Köszegi, "Is Addiction 'Rational'? Theory and Evidence," *The Quarterly Journal of Economics 116* (2001): 1261–1303.

Gruber, Jonathan, and Alan B. Krueger, "The Incidence of Mandated Employer Provided Health Insurance: Lessons from Workers' Compensation Insurance," in *Health Benefits and the Workforce*, David Bradford (ed.) (Washington, DC: U.S. Government Printing Office, 1992).

Gruber, Jonathan, and Brigitte C. Madrian, "Health Insurance, Labor Supply, and Job Mobility: A Critical Review of the Literature," *NBER* Working Paper 8817, March 2002.

Gruber, Jonathan, and Maria Owings, "Physician Financial Incentives and the Diffusion of Cesarean Section Delivery," *RAND Journal of Economics 27* (1996): 99–123.

Gruber, Jonathan, and Jonathan Zinman, "Youth Smoking in the U.S.: Evidence and Implications," *NBER* Working Paper 8872, 2002.

Grytten, Jostein, and Rune Sørensen, "Patient Choice and Access to Physician Services in Norway," *Health Economics, Policy and Law 4* (2008): 11–27.

Grytten, Jostein, and Rune Sørensen, "Primary Physician Services—List Size and Primary Physician Service Production," *Journal of Health Economics 26* (2007): 742–762.

Grytten, Jostein, and Rune Sørensen, "Type of Contract and Supplier-Induced Demand for Primary Physicians in Norway," *Journal of Health Economics 20* (2001): 379–393.

Grytten, Jostein, Irene Skau, and Rune Sørensen, "Do Expert Patients Get Better Treatment than Others? Agency Discrimination and Statistical Discrimination in Obstetrics," *Journal of Health Economics 30* (2011): 163–180.

Gulley, O. David, and Rexford E. Santerre, "The Effect of Tax Exemption on the Market Share of Nonprofit Hospitals," *National Tax Journal 46* (1993): 477–486.

Gulley, O. David, and Rexford E. Santerre, "Market Structure Elements: The Case of California Nursing Homes," *Journal of Health Care Finance 33* (2007): 1–16.

Haas-Wilson, Deborah, "Consumer Information and Providers' Reputations: An Empirical Test in the Market for Psychotherapy," *Journal of Health Economics 9* (1990): 321–333.

Hackmann, Martin B., Jonathan T. Kolstad, and Amanda E. Kowalski, "Adverse Selection and the Individual Mandate: When Theory Meets Practice," *American Economic Review 105* (2015): 1030–1066.

Hadley, Jack, *More Medical Care, Better Health* (Washington, DC: Urban Institute, 1982).

Hadley, Jack, "Medicare Spending and Mortality Rates of the Elderly," *Inquiry 25* (1988): 485–493.

Hadley, Jack, and Stephen Zuckerman, "The Role of Efficiency in Hospital Rate Setting," *Journal of Health Economics 13* (1994): 335–340.

Hall, Mark A., and Christopher J. Conover, "The Impact of Blue Cross Conversions on Accessibility, Affordability, and the Public Interest," *The Milbank Quarterly 81* (2003): 509–542.

Hall, Thomas D., and Cotton M. Lindsay, "Medical Schools: Producers of What? Sellers to Whom?" *Journal of Law and Economics 23* (1980): 55–80.

Hamilton, James L., "The Demand for Cigarettes: Advertising, the Health Scare, and the Cigarette Advertising Ban," *Review of Economics and Statistics 54* (1972): 401–411.

Hamilton, James L., "The Effect of Cigarette Advertising Bans on Cigarette Consumption," Proceedings of the Third World Conference on Smoking and Health (Washington, DC: U.S. DHEW, 1975).

Hanewinkel, Reiner, Christian Radden, and Tobias Rosenkranz, "Price Increase Causes Fewer Sales of Factory-Made Cigarettes and Higher Sales of Cheaper Loose Tobacco in Germany," *Health Economics 17* (2007): 683–693.

Hanratty, Maria, "Canadian National Health Insurance and Infant Health," *American Economic Review 86* (1996): 276–284.

Hansen, W. Lee, "Shortages and Investment in Health Manpower," in *The Economics of Health and Medical Care: Proceedings of the Conference* on *the Economics of Health and Medical Care*, May 10–12, 1962 (Ann Arbor, MI: University of Michigan, 1964).

Hansmann, Henry B., "The Effect of Tax Exemption and Other Factors on the Market Share of Nonprofit versus For-Profit Firms," *National Tax Journal 40* (1987): 71–82.

Hansmann, Henry B., "The Role of Nonprofit Enterprise," *Yale Law Journal 89* (1980): 835–901.

Hansmann, Henry, Daniel Kessler, and Mark McClellan, "Ownership Form and Trapped Capital in the Hospital Industry," *NBER* Working Paper 8989, 2002.

Harris, Jeffrey E., and Sandra W. Chan, "The Continuum-of-Addiction: Cigarette Smoking in Relation to Price among Americans," *Health Economics Letters 2* (1998): 3–12.

Harris, Jeffrey E., and Sandra W. Chan, "The Continuum-of-Addiction: Cigarette Smoking in Relation to Price among Americans Aged 15–29," *Health Economics 8* (1999): 81–86.

Harris, Jennifer L., et al., "A Crisis in the Marketplace: How Food Marketing Contributes to Childhood Obesity and What Can be Done," *Annual Review of Public Health 30* (2009): 211–225.

Harrison, Teresa, "Consolidation and Closures: An Empirical Analysis of Exits from the Hospital Industry," *Health Economics 16* (2007): 457–474.

Hassan, Mahmud, Gerard Wedig, and Michael A. Morrisey, "Charity Care by Non-Profit Hospitals: The Price of Tax-Exempt Debt," *International Journal of the Economics of Business 7* (2000): 47–62.

Hay, Joel W., and Gerard Anderson, "The Hospital Services Market: A Disequilibrium Analysis," *Southern Economic Journal 54* (1988): 656–665.

Hayek, Frederich, *The Constitution of Liberty* (Chicago: The University of Chicago Press, 1960).

Helland, Eric, and Mark Showalter, "The Impact of Liability on the Physician Labor Market," *Journal of Law and Economics 52* (2009): 635–663.

Henderson, Rebecca, and Iain Cockburn, "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery," *RAND Journal of Economics 27* (1996): 32–59.

Hill, Steven C., and Barbara L. Wolfe, "Testing the HMO Competitive Strategy: An Analysis of Its Impact on Medical Care Resources," *Journal of Health Economics 16* (1997): 261–286.

Himmelstein, David U., et al., "A Comparison of Hospital Administrative Costs in Eight Nations: U.S. Costs Exceed All Others by Far," *Health Affairs 33* (2014): 1586–1594.

Hirsch, Barry T., and Edwin J. Schumaker, "Classic or New Monopsony? Searching for Evidence in Nursing Labor Markets," *Journal of Health Economics 24* (2005): 969–989.

Hirsch, Barry T., and Edward Schumacher, "Monopsony Power and Relative Wages in the Labor Market for Nurses," *Journal of Health Economics 14* (1995): 443–476.

Hirsch, Barry T., and Edward Schumacher, "Union Wages, Rents, and Skills in Health Care Labor Markets," *Journal of Labor Research 19* (1998): 125–147.

Hirth, Richard A., "Consumer Information and Competition between Nonprofit and For-Profit Nursing Homes," *Journal of Health Economics 18* (1999): 219–240.

Hirth, Richard A., Michael E. Chernew, and Sean M. Orzol, "Ownership, Competition, and the Adoption of New Technologies and Cost Saving Practices in a Fixed-Price Environment," *Inquiry 37* (2000): 282–294.

Hirth, Richard A., et al., "Connecticut's Value-Based Insurance Plan Increased the Use of Targeted Services and Medication Adherence," *Health Affairs 35* (2016): 637–646.

Ho, Katherine, and Ariel Pakes, "Do Physician Incentives Affect Hospital Choice? A Progress Report," *International Journal of Industrial Organization 29* (2011): 317–322.

Hoadley, Jack et al., "The Medicare Part D Coverage Gap: Costs and Consequences in 2007," Kaiser Family Foundation, August 2007, www.kff.org/medicare/upload/7811.pdf, accessed June 7, 2011.

Hodgson, Thomas A., "Cigarette Smoking and Lifetime Medical Expenditures," *Milbank Quarterly 70* (1992): 81–125.

Hoff, Timothy J., "Doing the Same and Earning Less: Male and Female Physicians in a New Medical Specialty," *Inquiry 41* (2004): 301–315.

Hollingsworth, Bruce, "The Measurement of Efficiency and Productivity of Health Care Delivery," *Health Economics 17* (2008): 1107–1128.

Holtmann, Alphonse G., and Todd L. Idson, "Wage Determination of Registered Nurses in Proprietary and Nonprofit Nursing Homes," *Journal of Human Resources 28* (1993): 55–79.

Hornbrook, Mark C., and Alan C. Monheit, "The Contribution of Case-Mix Severity to the Hospital Cost-Output Relation," *Inquiry 22* (1985): 259–271.

Howard, David H., "Quality and Consumer Choice in Healthcare: Evidence from Kidney Transplantation," *Topics in Economic Analysis & Policy 5* (2005): Article 24.

Hsiao, William C., "Resource-Based Relative Value Scale: An Option for Physician Payment," *Inquiry 24* (1987): 360–361.

Hsiao, William C., Daniel L. Dunn, and Diana K. Verrilli, "Assessing the Implementation of Physician-Payment Reform," *New England Journal of Medicine 328* (1993): 928–933.

Hughes, David, and Alistair McGuire, "Stochastic Demand, Production Responses and Hospital Costs," *Journal of Health Economics 22* (2003): 999–1010.

Hughes, Susan L., et al., "Impact of Home Care on Hospital Days: A Meta Analysis," *Health Services Research 32* (1997): 415–432.

Hughes, Bethan, "2007 FDA Drug Approvals: A Year in Flux," *Nature Reviews Drug Discovery 7* (2008): 107–109.

Hurd, Michael D., and Kathleen McGarry, "Medical Insurance and the Use of Health Care Services by the Elderly," *Journal of Health Economics 16* (1997): 129–154.

Hurley, Jeremiah, "An Overview of the Normative Economics of the Health Sector," in *Handbook of Health Economics*, Volume I, A.J. Culyer and J.P. Newhouse (eds.) (Amsterdam: Elsevier Science B.V., 2000).

Hurley, Jeremiah, et al., "Publicly Funded Medical Savings Accounts: Expenditures and Distributional Impacts in Ontario, Canada," *Health Economics 17* (2008): 1129–1151.

Hwang, Wenke, et al., "Effects of Integrated Delivery System on Cost and Quality," *American Journal of Managed Care 19* (2013): e175–e184.

Illich, Ivan, *Medical Nemesis* (New York: Bantam Books, 1976).

Illife, Steve, and Sasha Shepperd, "What Do We Know about Hospital at Home: Lessons from International Experience," *Applied Health Economics and Health Policy 1* (2002): 141–147.

Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century* (Washington, DC: 2001).

Institute of Medicine, *To Err Is Human: Building a Safer Health System* (Washington, DC, 1999).

International Agency for Research on Cancer, Non-Ionizing Radiation, Part 2: Radiofrequency Electromagnetic Fields 102 (2013), http://monographs.iarc.fr/ENG/Monographs/vol102/index.php, accessed July 1, 2016.

International Health, *How Australia Compares*, www.aihw.gov.au/publications/health/ihhac, 2002, accessed November 2016.

Islam, M. Kamrul, et al., "Social Capital and Health: Does Egalitarianism Matter? A Literature Survey," *International Journal for Equity in Health 5* (2006): 1–28.

Islam, M. Kamrul et al., "Social Externalities and Mortality in Sweden," *Economics and Human Biology 6* (2008): 19–42.

Iversen, Tor (2004), "The Effects of a Patient Shortage on General Practitioners Future Income and Patient Lists," *Journal of Health Economics 23* (2004): 673–694.

Iwasaki, Natsuko, Carol H. Tremblay, and Victor J. Tremblay, "Advertising Restrictions and Cigarette Smoking: Evidence from Myopic and Rational Addiction Models," *Contemporary Economic Policy 24* (2006): 370–381.

Jacob, Paul D., Noelia Duchovny, and Brandy J. Lipton, "Changes in Health Status and Care Use after ACA Expansions among the Insured and Uninsured," *Health Affairs 35* (2016): 1184–1188.

Jacobson, Gretchen, Anthony Damico, Tricia Neuman, and Marsha Gold, "Medicare Advantage 2015 Spotlight: Enrollment Market Update," Kaiser Family Foundation, June 2015, Figure 5, http://files.kff.org/attachment/issue-brief-medicare-advantage-2015-spotlight-enrollment-market-update, downloaded July 10, 2016.

Jacobson, Peter D., Louise E. Parker, and Ian D. Coulter, "Nurse Practitioners and Physician Assistants as Primary Care Providers in Institutional Settings," *Inquiry 35* (1998/99): 432–446.

Jaegher, Kris De, and Marc Jegers, "A Model of the Physician with Demand Inducement," *Journal of Health Economics 19* (2000): 231–258.

Jantzen, Robert, and Patricia R. Loubeau, "Hospital Selection by Managed Care Insurers," *Health Care Management Review 25* (2000): 93–102.

Jayachandran, Seema, et al., "Modern Medicine and the Twentieth Century Decline in Mortality: Evidence on the Impact of Sulfa Drugs," *American Economic Journal: Applied Economics 2* (2010): 118–146.

Jensen, Gail A., and Allen C. Goodman, "The Economics of Gynecology and Obstetrics," in *Practical Strategies in Obstetrics and Gynecology*, Scott B. Ransom (ed.) (New York: W. B. Saunders, 1999): 719–733.

Jensen, Gail A., and Michael A. Morrisey, "Endogenous Fringe Benefits, Compensating Wage Differentials and Older Workers," *International Journal of Health Care Finance and Economics 1* (2001): 203–226.

Jensen, Gail A., and Michael A. Morrisey, "The Role of Physicians in Hospital Production," *Review of Economics and Statistics 68* (1986): 432–442.

Jha, Ashish K., et al., "Care in U.S. Hospitals—The Hospital Quality Alliance Program," *New England Journal of Medicine 353* (2005): 265–274.

Jiang, H. Joanna, Bernard Friedman, and Shenyi Jiang, "Hospital Cost and Quality Performance in Relation to Market Forces," *Health Care Finance and Economics 13* (2013): 53–71.

Jiménez-Martin, Serji, José M. Labeaga, and Maite Martinez-Granado, "An Empirical Analysis of the Demand for Physician Services across the European Union," *European Journal of Health Economics 5* (2004): 150–165.

Joesch, Jutta M., Thomas M. Wickizer, and Paul J. Feldstein, "Does Competition by Health Maintenance Organizations Affect the Adoption of Cost-Containment Measures by Fee-for-Service Plans?" *American Journal of Managed Care 4* (1998): 832–838.

Johansen, Michael E., and Caroline Richardson, "Estimation of Potential Savings through Therapeutic Substitution," *Journal of the American Medical Association 176* (2016): 769–775.

Johnston, Jack, "Advertising and the Aggregate Demand for Cigarettes: A Comment," *European Economic Review 14* (1980): 117–125.

Jones, Andrew M., "Smoking Cessation and Health: A Response," *Journal of Health Economics 15* (1996): 755–759.

Jones, Robert F., et al., "Review of U.S. Medical School Finances, 1996–1997," *Journal of the American Medical Association 280* (1998): 813–818.

Joyce, Geoffrey F., et al., "Employer Drug Benefit Plans and Spending on Prescription Drugs," *Journal of the American Medical Association 288* (2002): 1733–1739.

Joyce, Theodore, "Impact of Augmented Prenatal Care on Birth Outcomes of Medicaid Recipients in New York City," *Journal of Health Economics 18* (1999): 31–67.

Joyce, Theodore, "What Can Massachusetts Teach Us about National Health Insurance Reform?" *Journal of Policy Analysis and Management 30* (2011): 177–195.

Joyce, Theodore, Andrew D. Racine, and Naci Mocan, "The Consequences and Costs of Maternal Substance Abuse in New York City: A Pooled Time-Series, Cross-Section Analysis," *Journal of Health Economics 11* (1992): 297–314.

Joynt, Karen E., et al., "Hospital Closures Had No Measurable Impact on Local Hospitalization Rates or Mortality Rates, 2003–11," *Health Affairs 34* (2015): 765–772.

Jung, Hye-Young, et al., "Integrated Medicare and Medicaid Managed Care and Rehospitalization of Dual Eligibles," *American Journal of Managed Care 21* (2015): 711–719.

Jung, Ki-Taig, "Influence of the Introduction of a Per-Visit Copayment on Health Care Use and Expenditures: The Korean Experience," *Journal of Risk and Insurance 65* (1998): 35–56.

Kaestner, Robert, and Nasreen Kahn, "Medicare Part D and Its Effect on the Use of Prescription Drugs and Use of Other Health Care Services of the Elderly," *Journal of Policy Analysis and Management 31* (2012): 253–279.

Kaestner, Robert, Theodore Joyce, and Hassan Wehbeh, "The Effect of Maternal Drug Use on Birth Weight: Measurement Error in Binary Variables," *Economic Inquiry 34* (October 1996): 617–629.

Kaestner, Robert, Cuiping Long, and G. Caleb Alexander, "Effects of Prescription Drug Insurance on Hospitalization and Mortality: Evidence from Medicare Part D," September 2014 NBER Working Paper 19948.

Kaiser Family Foundation, *Employer Health Benefits Survey*, September 2015, Washington DC, http://kff.org/health-costs/report/2015-employer-health-benefits-survey/, accessed November 2016.

Kaiser Family Foundation, "Summary of New Health Reform Law," April 15, 2011. Available at www.kff.org/healthreform/upload/8061.pdf, accessed February 22, 2017.

Kaiser Family Foundation, *Trends and Indicators in the Changing Health Care Marketplace*, various years.

Kalist, David E., "Registered Nurses and the Value of Bilingualism," *Industrial & Labor Relations Review 59* (2005): 101–118.

Kantarevic, Jasmin, Boris Kralj, and Darrel Weinkauf, "Enhanced Fee-for-Service Model and Physician Productivity: Evidence from Family Health Groups in Ontario," *Journal of Health Economics 30* (2011): 99–111.

Kaplan, Robert M., and John P. Anderson, "A General Health Policy Model: Update and Applications," *Health Services Research 23* (1988): 203–235.

Kapur, Kanika, Carole Roan Gresenz, and David M. Studdert, "Managing Care: Utilization Review in Action at Two Capitated Medical Groups Health Affairs," 10.1377/hlthaff.w3.275, posted June 18, 2003, accessed June 9, 2008.

Katzman, Brett, Sara Markowitz, and Kerry Anne McGeary, "An Empirical Investigation of the Social Market for Cigarettes," *Health Economics 16* (2007): 1025–1039.

Kawachi, Ichiro, "Social Capital and Community Effects in Population and Individuals," *Annals of the New York Academy of Sciences 896* (1999): 120–130.

Keehan, Sean P., et al., "National Health Expenditure Projections, 2014–24: Spending Growth Faster than Recent Trends," *Health Affairs 34* (2015): 1407–1417.

Keeler, Emmett B., and John E. Rolph, "The Demand for Episodes of Treatment in the Health Insurance Experiment," *Journal of Health Economics 7* (1988): 333–367.

Keeler, Emmett B., Willard G. Manning, Jr., and Kenneth B. Wells, "The Demand for Episodes of Mental Health Services," *Journal of Health Economics 7* (1988): 369–392.

Keeler, Theodore E., et al., "Taxation, Regulation, and Addiction: A Demand Function for Cigarettes Based on Time-Series Evidence," *Journal of Health Economics 12* (1993): 1–18.

Keeler, Theodore E., et al., "The U.S. National Tobacco Settlement: The Effects of Advertising and Price Changes on Cigarette Consumption," *Applied Economics 36* (2004): 1623–1629.

Kenkel, Donald S., "The Demand for Preventive Medical Care," *Applied Economics 26* (1994): 313–325.

Kenworthy, Lane, "It's Hard to Make It in America: How America Stopped Being the Land of Opportunity," *Foreign Affairs 91* (2012): 103–109.

Kessel, Reuben, "Price Discrimination in Medicine," *Journal of Law and Economics 1* (1958): 20–53.

Kessler, Daniel P., and Mark B. McClellan, "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics 115* (2000): 577–615.

Kim, Daniel, and Ichiro Kawachi, "Food Taxation and Pricing Strategies to 'Thin Out' the Obesity Epidemic," *American Journal of Preventive Medicine 30* (2006): 430–437.

Kochi, I., B. Hubbell, and R. Kramer., "An Empirical Bayes Approach to Combining and Comparing Estimates of the Value of a Statistical Life for Environmental Policy Analysis," *Environmental and Resource Economics 34* (2006): 385–406.

Konetzka, R. Tamara, and Ye Luo, "Explaining Lapse in Long-Term Care Insurance Markets," *Health Economics 20* (2010): 1169–1183.

Koolman, Xander, and Eddy van Doorslaer, "On the Interpretation of a Concentration Index of Inequality," *Health Economics 13* (2004): 649–656.

Kooreman, Peter, "Data Envelopment Analysis and Parametric Frontier Estimation: Complementary Tools," *Journal of Health Economics 13* (1994): 345–346.

Kooreman, Peter, "Nursing Home Care in the Netherlands: A Nonparametric Efficiency Analysis," *Journal of Health Economics 13* (1994): 301–316.

Kowalski, Amanda E., "Estimating the Tradeoff between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance," *International Journal of Industrial Organization 43* (2015): 122–135.

Krueger, Alan B., and Uwe E. Reinhardt, "The Economics of Employer versus Individual Mandates," *Health Affairs 13* (1994): 34–53.

Kseib, Nicolas, Javier Urzay, and Gianluca Iaccarino, "Statistical Inference of Uncertainties in Elementary Reaction Rates of Chemical Mechanisms," Center for Turbulence Research Annual Research Briefs 2011, http://web.stanford.edu/~jurzay/Urzay2011CTR3.pdf, accessed January 13, 2016.

Ku, Leighton, "Medicaid: Improving Health, Saving Lives," Center on Budget and Policy Priorities Analysis of National Health Interview Survey data, www.kff.org/medicaid/upload/2177-05.pdf, accessed August 2005.

Labelle, Roberta, Gregg Stoddard, and Thomas Rice, "A Re-Examination of the Meaning and Importance of Supplier-Induced Demand," *Journal of Health Economics 13* (1994): 347–368.

Laine, Juha, et al., "Measuring the Productive Efficiency and Clinical Quality of Institutional Long-Term Care for the Elderly," *Health Economics 14* (2005): 245–256.

Lakdawalla, Darius, and Tomas Philipson, "The Nonprofit Sector and Industry Performance," *Journal of Public Economics 90* (2006): 1681–1698.

Lalonde, Marc, *A New Perspective on the Health of Canadians*, Office of the Canadian Minister of National Health and Welfare, Ottawa, 1974.

Lamberton, C.E., W.D. Ellington, and K.R. Spear, "Factors Determining the Demand for Nursing Home Services," *Quarterly Review of Economics and Business 26* (1986): 74–90.

Landon, Bruce E., et al., "Quality Monitoring and Management in Commercial Health Plans," *American Journal of Managed Care 14* (2008): 377–386.

Lanoie, Paul, and Paul Leclair, "Taxation or Regulation: Looking for a Good Anti-Smoking Policy," *Economics Letters 58* (1998): 85–89.

Laugesen, Murray, and Chris Meads, "Tobacco Advertising Restrictions, Price, Income, and Tobacco Consumption in OECD Countries, 1960–1986," *British Journal of Addiction 86* (1991): 1343–1354.

Law, Michael R., and Karen A. Grépin, "Is Newer Always Better? Re-Evaluating the Benefits of Newer Pharmaceuticals," *Journal of Health Economics 29* (2010): 743–750.

Lee, Dwight R., "Why Workers Should Want Mandated Benefits to Lower Their Wages," *Economic Inquiry 34* (1996): 401–407.

Lee, Joy Li-Yueh, et al., "A Systematic Review of Reference Pricing: Implications for U.S. Prescription Drug Spending," *American Journal of Managed Care 18* (2012): e429–e437.

Lee, Joy L., et al., "Value-Based Insurance Design: Quality Improvement but No Cost Savings," *Health Affairs 32* (2013): 1251–1257.

Lee, Robert H., and Jack Hadley, "Physicians' Fees and Public Medical Care Programs," *Health Services Research 16* (1981): 185–203.

Lee, Roger I., and Lewis W. Jones, *The Fundamentals of Good Medical Care*, publication of the Committee on the Costs of Medical Care (Chicago: University of Chicago, 1933).

Lee, Young Hoon, and Peter Schmidt, "A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency," in *Measurement of Production Efficiency*, Harold O. Fried, C. A. Knox Lovell, et al. (eds.) (New York: Oxford University Press, 1993).

Leffler, Keith B., "Persuasion or Information? The Economics of Prescription Drug Advertising," *Journal of Law and Economics 24* (1981): 45–74.

Leffler, Keith B., "Physician Licensure: Competition and Monopoly in American Medicine," *Journal of Law and Economics 21* (1978): 165–186.

Lehner, Laura A., and James F. Burgess, "Teaching and Hospital Production: The Use of Regression Estimates," *Health Economics 4* (1995): 113–125.

Leibowitz, Arlene A., "The Demand for Health and Health Concerns after 30 Years," *Journal of Health Economics 23* (2004): 663–671.

Leibowitz, Arlene A., "An Economic Perspective on Work, Family and Well-Being," in S. M. Bianchi, L. M. Casper, and R. B. King (eds.), *Work, Family, Health, and Well-Being* (Mahwah, New Jersey: Lawrence Erlbaum, 2005).

Leigh, J. Paul, et al., "Lifetime Earnings for Physicians across Specialties," *Medical Care 50* (2012): 1093–1101.

Leonard, Kenneth L., "Is Patient Satisfaction Sensitive to Changes in the Quality of Care? An Exploitation of the Hawthorne Effect," *Journal of Health Economics 27* (2008): 444–459.

Leonard, Kenneth L., and Joshua G. Zivin, "Outcome versus Service Based Payments in Health Care: Lessons from African Traditional Healers," *Health Economics 14* (2005): 575–593.

Lerner, Abba P., "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies 11* (1943): 157–175.

Levit, Katherine, et al., "Health Spending Rebound Continues in 2002," *Health Affairs 23* (2004): 147–150.

Levitt, Steven D., and Jack Ray Porter, "Estimating the Effect of Alcohol on Driver Risk Using Only Fatal Accident Statistics," *NBER* Working Paper 6944, 1999.

Levy, Helen, and David Weir, "Take-Up of Medicare Part D: Results from the Health and Retirement Study," *Journals of Gerontology B Psychology Science Social Science 65B* (2010): 492–501.

Lewit, Eugene M., and Douglas Coate, "The Potential for Using Excise Taxes to Reduce Smoking," *Journal of Health Economics 1* (1982): 121–145.

Lewit, Eugene M., Douglas Coate, and Michael Grossman, "The Effects of Government Regulation on Teenage Smoking," *Journal of Law and Economics 24* (1981): 545–569.

Li, Tong, and Robert Rosenman, "Estimating Hospital Costs with a Generalized Leontief Function," *Health Economics 10* (2001): 523–538.

Liang, J. Nellie, and Jonathon D. Ogur, *Restrictions on Dental Auxiliaries: An Economic Policy Analysis* (Washington, DC: Federal Trade Commission, 1987).

Lichtenberg, Frank R., "Benefits and Costs of New Drugs: An Update," *Managerial and Decision Economics 28* (2007a): 485–490.

Lichtenberg, Frank R., "Have Newer Cardiovascular Drugs Reduced Hospitalization? Evidence from Longitudinal Country-Level Data on 20 OECD Countries, 1995–2003," *Health Economics 18* (2009): 519–534.

Lichtenberg, Frank R, "Importation and Innovation," *Economics of Innovation and New Technology 16* (2007b): 403–417.
Lichtenberg, Frank R., "Pharmaceutical Innovation, Mortality Reduction, and Economic Growth," *NBER* Working Paper 6569, 2006.
Lien, Hsien-Ming, et al., "Hospital Ownership and Performance: Evidence from Stroke and Cardiac Treatment in Taiwan," *Journal of Health Economics 27* (2008): 1208–1223.
Light, Donald W., and Rebecca Warburton, "Demythologizing the High Costs of Pharmaceutical Research," *BioSocieties 6* (2011): 34–50.
Light, Donald W., and Rebecca N. Warburton, "Extraordinary Claims Require Extraordinary Evidence," *Journal of Health Economics 24* (2005): 1030–1033.
Lin, Haizen, and Jeffrey Prince, "The Impact of the Partnership Long-Term Care Insurance Program on Private Coverage," *Journal of Health Economics 32* (2013): 1205–1213.
Lindeboom, Maarten, Ana Llena-Nozai, and Bas van der Klaauw, "Parental Education and Child Health: Evidence from a Schooling Reform," Tinbergen Institute, Discussion Paper No. 2516 (2006).
Lindhjem, Henrik., S. Navrud, and N.A. Braathen, "Valuing Lives Saved from Environmental Transport and Health Policies: A Meta-Analysis of Stated Preference Studies, Prepared for the Organization for Economic Co-operation and Development," ENV/EPOC/WPNEP(2008)10/FINAL, (2010).
Lindrooth, Richard C., Anthony T. Lo Sasso, and Gloria J. Bazzoli, "The Effect of Urban Hospital Closure on Markets," *Journal of Health Economics 22* (2003): 691–712.
Lindstrom, Martin, and Mohabbat Moseni, "Social Capital, Trust, and Self-Reported Psychological Health: A Population Based Study," *Social Science & Medicine 68* (2009): 436–433.
Linna, Miika, and Unto Häkkinen, "Reimbursing for Costs of Teaching and Research in Finnish Hospitals: A Stochastic Frontier Approach," *International Journal of Health Care Finance and Economics 6* (2006): 83–97.
Liu, Liqun, Andrew J. Rettenmaier, and Thomas R. Saving, "Longevity Bias in Cost-Effectiveness Analysis," *Health Economics 17* (2008): 523–534.
Liu, Zhimei, William H. Dow, and Edward C. Norton, "Effect of Drive-Through Delivery Laws on Postpartum Length of Stay and Hospital Charges," *Journal of Health Economics 23* (2004): 129–155.
Lleras-Muney, Adriana, "The Relationship between Education and Adult Mortality in the United States," *Review of Economic Studies 72* (2005): 189–221.
Lo Sasso, Anthony T., et al., "The $16,819 Pay Gap for Newly Trained Physicians: The Unexplained Trend of Men Earning More than Women," *Health Affairs 30* (2011): 193–200.
Locke, John, *Epistola de Tolerantia, John Locke Bibliography*, www.libraries.psu.edu/iasweb/locke/ch0b.htm, July 1, 2002.
Locke, John, *On the Poor Law and Working Schools* (1697) eco.utexas.edu/~hmcleave/350kPEELockePoorEd Table.pdf, accessed May 14, 2006.
Locke, John, *Two Treatises of Government* (1690; Cambridge: Cambridge University Press, 1960).
Lohr, Steve, "Most Doctors Aren't Using Electronic Records," *New York Times*, June 19, 2008, www.nytimes.com/2008/06/19/technology/19patient.html, accessed November 2016.
Long, Stephen H., Russell F. Settle, and Charles W. Wrightson Jr., "Employee Premiums, Availability of Alternative Plans, and HMO Disenrollment," *Medical Care 26* (1988): 927–938.
Loury, Glenn, "A Dynamic Theory of Racial Differences," in *Women, Minorities and Employment Discrimination*, Phyllis B. Wallace and Annette LaMond (eds.) (Lexington, MA: Lexington Books, 1977).

Lovell, C. A. Knox, and Shelton S. Schmidt (eds.), *The Measurement of Productive Efficiency* (Oxford: Oxford University Press, 1993).

Lu, John Z., and William S. Comanor, "Strategic Pricing of New Pharmaceuticals," *Review of Economics and Statistics 80* (1998): 108–117.

Luft, Harold S., "Health Maintenance Organizations and the Rationing of Medical Care," *Milbank Memorial Fund Quarterly/Health and Society 60* (1982): 268–306.

Luft, Harold S., *Health Maintenance Organizations: Dimensions of Performance* (New York: John Wiley & Sons, 1981).

Luft, Harold S., "How Do Health-Maintenance Organizations Achieve Their 'Savings'?" *New England Journal of Medicine 298* (1978): 1336–1343.

Lundborg, Petter, "Social Capital and Substance Abuse among Swedish Adolescents," *Social Science & Medicine 61* (2005): 1151–1158.

Lundborg, Petter, and Bjorn Lindren, "Do They Know What They Are Doing? Risk Perception and Smoking Behavior among Swedish Teenagers," *Journal of Risk and Uncertainty 28* (2004): 261–286.

MacDonald, Ziggy, "What Price Drug Use? The Contribution of Economics to an Evidence-Based Drugs Policy," *Journal of Economic Surveys 18* (2004): 113–152.

Maciosek, Michael V., et al., "Greater Use of Preventive Services in U.S. Health Care Could Save Lives at Little or No Cost," *Health Affairs 29* (2010): 1656–1660.

Madrian, Brigitte C., "Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock?" *Quarterly Journal of Economics 109* (1994): 27–54.

Magnusson, Jon, "Efficiency Measurement and the Operationalization of Hospital Production," *Health Services Research 31* (1996): 21–37.

Mahmoudi, Elham, and Gail A. Jensen, "Diverging Racial and Ethnic Disparities in Access to Physician Care," *Medical Care 50* (2012): 327–334.

Mandon, Bruce E., et al., "Quality Monitoring and Management in Commercial Health Plans," *American Journal of Managed Care 14* (2008): 377–386.

Manning, Willard G., and M. Susan Marquis, "Health Insurance: The Tradeoff between Risk Pooling and Moral Hazard," *Journal of Health Economics 15* (1996): 609–639.

Manning, Willard G., et al., "A Controlled Trial of the Effect of a Prepaid Group Practice on Use of Services," *New England Journal of Medicine 310* (1984): 1505–1510.

Manning, Willard G., et al., *The Costs of Poor Health Habits* (Cambridge, MA: Harvard University Press, 1991).

Manning, Willard G. et al., "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review 77* (1987): 251–277.

Mark, Tami L., "Analysis of the Rationale for, and Consequences of, Nonprofit and For-Profit Ownership Conversions," *Health Services Research 34* (1999): 83–101.

Mark, Tami L., "Psychiatric Hospital Ownership and Performance: Do Nonprofit Organizations Offer Advantages in Markets Characterized by Asymmetric Information?" *Journal of Human Resources 31* (1996): 631–649.

Marquis, M. Susan, Jeannette A. Rogowski, and Jose Escarcé, "The Managed Care Backlash: Did Consumers Vote with Their Feet?" *Inquiry 41* (2004/2005): 376–390.

Martinez, Barbara, and Jacob Goldstein, "Big Pharma Faces Grim Prognosis," *Wall Street Journal*, December 6, 2007, p. 1.

Marteau, Theresa, et al., "Downsizing: Policy Options to Reduce Portion Sizes to Help Tackle Obesity," *British Medical Journal 351* (2015): h5863.

Martin, Anne B., et al., "National Health Spending in 2014: Faster Growth Driven by Coverage Expansion and Prescription Drug Spending," *Health Affairs 35* (2016): 150–160.

Matawie, Kenan, and Albert Assaf, "Bayesian and DEA Efficiency Modeling: An Application to Hospital Foodservice Operations," *Journal of Applied Statistics 37* (2010): 945–953.

Mathauer, Inke, and Friedrich Wittenbecher, *DRG-Based Payment Systems in Low- and Middle Income Countries: Implementation Experiences and Challenges,*" World Health Organization, Geneva, 2012.

Mather, Dell, B., "Incorporating Clinical Outcomes and Economic Consequences into Drug Formulary Decisions: A Practical Approach," *American Journal of Managed Care 5* (1999): 277–285.

Mays, Glen P., and Sharla A. Smith, "Evidence Links Increases in Public Health Spending to Declines in Preventable Deaths," *Health Affairs 30* (2011): 1585–1593.

McCarthy, Patrick S., "The Effect of Higher Rural Interstate Speed Limits in Alcohol-Related Accidents," *Journal of Health Economics 12* (1993): 281–299.

McCarthy, Thomas R., "The Competitive Nature of the Primary-Care Physician Services Market," *Journal of Health Economics 4* (1985): 93–117.

McClellan, Mark, "Hospital Reimbursement Incentives: An Empirical Analysis," *Journal of Economics and Management Strategy 6* (1997): 91–128.

McClellan, Mark, and Douglas Staiger, "Comparing Hospital Quality at For-Profit and Not-For-Profit Hospitals," *NBER* Working Paper 7324, 1999.

McCullough, Jeffrey S., "The Adoption of Hospital Information Systems," *Health Economics 17* (2008): 649–664.

McDevitt, Roland, et al., "Risk Selection into Consumer-Directed Health Plans: An Analysis of Family Choices," *Health Services Research 49* (2014): 609–627.

McGlynn, Elizabeth A., et al., "The Quality of Health Care Delivered to Adults in the United States," *New England Journal of Medicine 348* (2003): 2635–2645.

McGuinness, Tony, and Keith Cowling, "Advertising and the Aggregate Demand for Cigarettes," *European Economic Review 6* (1975): 311–328.

McGuire, Thomas G., and Mark V. Pauly, "Physician Response to Fee Changes with Multiple Payers," *Journal of Health Economics 10* (1991): 385–410.

McKay, Niccie L., "The Prisoner's Dilemma: An Obstacle to Cooperation in Health Care Markets," *Medical Care Review 51* (1994): 179–204.

McKeown, Thomas, *The Modern Rise of Population* (New York: Academic Press, 1976).

McKinlay, John B., and Sonja M. McKinlay, "The Questionable Contribution of Medical Measures to the Decline of Mortality in the United States in the Twentieth Century," *Milbank Memorial Fund Quarterly/Health and Society 55* (1977): 405–428.

McLaughlin, Catherine G., "Market Responses to HMOs: Price Competition or Rivalry?" *Inquiry 25* (1988): 207–218.

McPherson, Klim, et al., "Small-Area Variations in the Use of Common Surgical Procedures: An International Comparison of New England, England, and Norway," *New England Journal of Medicine 307* (1982): 1310–1313.

Medicare Payment Advisory Commision (MedPAC), *Healthcare Spending and the Medicare Program*, June 2010, www.medpac.gov/documents/jun10datbookentirereport.pdf, accessed November 2016.

Mello, Michelle M., Amitabh Chandra, Atul A. Gawande, and David M. Studdert, "National Costs of the Medical Liability System," *Health Affairs 29* (2010): 1–9.

Mellor, Jennifer M., "Do Cigarette Taxes Affect Children's Body Mass Index? The Effect of Household Environment on Health," *Health Economics 20* (2011): 417–431.

Melnick, Glen, and Emmett Keeler, "The Effects of Multi-Hospital Systems on Hospital Prices," *Journal of Health Economics 26* (2007): 400–413.

Melnick, Glen A., et al., "The Effects of Market Structure and Bargaining Position on Hospital Prices," *Journal of Health Economics 11* (1992): 217–233.

Meng, Qingyue, et al., "People's Republic of China Health System Review," *Health Systems in Transition*, Vol. 5, No. 7 (2015), Asia Pacific Observatory on Health Systems and Policies, Geneva: World Health Organization.

Meyerhoeffer, Chad D., and Samuel H. Zuvekas, "New Estimates of the Demand for Physical and Mental Health Treatment," *Health Economics 19* (2010): 297–315.

Miller, Douglas L., et al., "Social Capital and Health in Indonesia," *World Development 34* (2006): 1084–1098.

Miller, Rebecca S., Harry S. Jonas, and Michael E. Whitcomb, "The Initial Employment Status of Physicians Completing Training in 1994," *Journal of the American Medical Association 275* (1996): 708–712.

Miller, Robert H., and Harold S. Luft, "Does Managed Care Lead to Better or Worse Quality of Care?" *Health Affairs 16* (1997): 7–25.

Miller, Robert H., and Harold S. Luft, "HMO Plan Performance Update: An Analysis of the Literature: 1997–2001," *Health Affairs 21* (2002): 63–86.

Miller, Robert H., and Harold S. Luft, "Managed Care Plan Performance Since 1980: A Literature Analysis," *Journal of the American Medical Association 271* (May 18, 1994): 1512–1519.

Miller, Theodore R., "Variation between Countries in Values of Statistical Life," *Journal of Transport Economics and Safety 34* (2000): 169–188.

Milne, C-P., and K.I. Kaitin, "Impact of the New U.S. Health-Care Reform Legislation on the Pharmaceutical Industry: Who Are the Real Winners?" *Clinical Pharmacology & Therapeutics 88* (2010): 589–592.

Min, Hery, "Large-Sized Soda Ban as an Alternative to Soda Tax," *Cornell Journal of Law and Public Policy 23* (2013), 187–232.

Miron, Jeffrey A., "The Effect of Alcohol Prohibition on Alcohol Consumption," *NBER* Working Paper 7130, 1999.

Mitchell, Janet B., and Jerry Cromwell, "Impact of All-or-Nothing Assignment Requirement under Medicare," *Health Care Financing Review 4* (1983): 59–78.

Mitchell, Janet B., and Jerry Cromwell, "Physician Behavior under the Medicare Assignment Option," *Journal of Health Economics 1* (1982): 245–264.

Mooney, Gavin et al., "Reweighing Heat: Response to Culyer, van Doorslaer, and Wagstaff," *Journal of Health Economics 11* (1992): 199–205.

Mooney, Gavin, et al., "Utilisation as a Measure of Equity: Weighing Heat?" *Journal of Health Economics 10* (1991): 475–480.

Moore, Michael J., "Death and Tobacco Taxes," *NBER* Working Paper 5153, 1995.

Morgan, Eleanor J., "Innovation and Merger Decisions in the Pharmaceutical Industry," *Review of Industrial Organization 19* (2001): 181–197.

Morrisey, Michael A., "Competition in Hospital and Health Insurance Markets: A Review and Research Agenda," *Health Services Research 36* (2001): 191–222.

Morrisey, Michael A., *Health Insurance*, Chapter 8 (Chicago: Health Administration Press, 2008).

Morrisey, Michael A., "Movies and Myths: Hospital Cost Shifting," *Business Economics 30* (1995): 22–25.

Moscone, Francesco, and Elisa Tosetti, "Health Expenditure and Income in the United States," *Health Economics 19* (2010): 1385–1403.

Mossialos, Elias, et al., *2015 International Profiles of Health Care Systems*, The Commonwealth Fund, January 2016, X-1857_Mossialos_intl_profiles_2015_v7.pdf, www.

commonwealthfund.org/~/media/files/publications/fund-report/2016/jan/1857_mossialos_intl_profiles_2015_v7.pdf, last accessed June 16, 2016.

Motheral, Brenda R., and Rochelle Henderson, "The Effect of a Copay Increase on Pharmaceutical Utilization, Expenditures, and Treatment Continuation," *American Journal of Managed Care 5* (1999): 1383–1394.

Mrozek, Janusz R., and Laura O. Taylor, "What Determines the Value of Life? A Meta-Analysis," *Journal of Policy Analysis and Management 21* (2002): 253–270.

Mueller, Curt D., and Alan C. Monheit, "Insurance Coverage and the Demand for Dental Care," *Journal of Health Economics 7* (1988): 59–72.

Mukamel, Dana B., and William D. Spector, "The Competitive Nature of the Nursing Home Industry: Price Mark Ups and Demand Elasticities," *Applied Economics 34* (2002): 413–420.

Mullahy, John, and Jody L. Sindelar, "Alcoholism, Work, and Income," *Journal of Labor Economics 11* (1993): 494–520.

Mullahy, John, and Jody Sindelar, "Do Drinkers Know When to Say When? An Empirical Analysis of Drunk Driving," *Economic Inquiry 22* (1994): 383–394.

Mullen, Kathleen J., Richard G. Frank, and Meredith B. Rosenthal, "Can You Get What You Pay for? Pay-for-Performance and the Quality of Healthcare Provider," *RAND Journal of Economics 41* (2010): 64–91.

Murphy, Kevin M., and Robert H. Topel, "The Value of Health and Longevity," *NBER* Working Paper 11405, June 2005.

Mushkin, Selma J., *Biomedical Research: Costs and Benefits* (Cambridge, MA: Ballinger, 1979).

Nassiri, Abdelhak, and Lise Riochaix, "Revisiting Physician Financial Incentives in Quebec: A Panel System," *Health Economics 15* (2006): 49–64.

National Association of Home Care, "Basic Statistics about Hospice," www.nahc.org/consumer/hpcstats, 1999.

Nelson, Phillip, "Advertising as Information," *Journal of Political Economy 82* (1974): 729–754.

Nelson, Phillip, "Information and Consumer Behavior," *Journal of Political Economy 78* (1970): 311–329.

Neumann, Peter J., Joshua T. Cohen, and Milton C. Weinstein, "Updating Cost-Effectiveness: The Curious Resilience of the $50,000-per-QALY Threshold," *New England Journal of Medicine 371* (2014): 796–797.

Newbold, Paul, William L. Carlson, and Betty Thorne, *Statistics for Business and Economics*, 6th ed. (Upper Saddle River, NJ: Prentice Hall, 2007).

Newhouse, Joseph P., *The Economics of Medical Care* (Reading, MA: Addison-Wesley, 1978).

Newhouse, Joseph P., "Frontier Estimation: How Useful a Tool in Health Economics?" *Journal of Health Economics 13* (1994): 317–322.

Newhouse, Joseph P., "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives 6* (1992): 3–21.

Newhouse, Joseph P., "Medical-Care Expenditures: A Cross-National Survey," *Journal of Human Resources 12* (1977): 115–125.

Newhouse, Joseph P., "Medicare Policy in the 1990s," *NBER* Working Paper 8531, 2001.

Newhouse, Joseph P., "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *American Economic Review 60* (1970): 64–74.

Newhouse, Joseph P., "Why Is There a Quality Chasm?" *Health Affairs 21* (2002): 13–25.

Newhouse, Joseph P., and Charles E. Phelps, "New Estimates of Price and Income Elasticities of Medical Care Services," in *The Role of Health Insurance in the Health Services Sector*, Richard Rosett (ed.) (New York: Neal Watson, 1976).

Newhouse, Joseph P., Geoffrey Anderson, and Leslie L. Roos, "Are Fee-for-Service Costs Increasing Faster than HMO Costs?" *Medical Care 23* (1985): 960–966.

Newhouse, Joseph P., Geoffrey Anderson, and Leslie L. Roos, "Hospital Spending in the United States and Canada: A Comparison," *Health Affairs 7* (1988): 6–16.

Newhouse, Joseph P., et al., *Free for All? Lessons from the RAND Health Insurance Experiment* (Cambridge, MA: Harvard University Press, 1993).

Nguyen, Nguyen X., and Frederick W. Derrick, "Physician Behavioral Response to a Medicare Price Reduction," *Health Services Research 32* (1997): 283–298.

Nicholson, Sean, "How Much Do Medical Students Know about Physician Income?" *Journal of Human Resources 40* (2005): 100–114.

Nord, Erik, "Health State Values from Multi-Attribute Utility Instruments Need Correction," *Annals of Medicine 33* (2001): 371–374.

Nordhaus, William, "Critical Assumptions in the Stern Review of Climate Change," *Science 317* (2007): 201–202.

Norton, Edward C., "Long-Term Care," in *Handbook of Health Economics*, Volume 1B, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam: Elsevier Science B.V., 2000).

Norton, Edward C., and Douglas O. Staiger, "How Hospital Ownership Affects Access to Care for the Uninsured," *RAND Journal of Economics 25* (1994): 171–185.

Nozick, Robert, *Anarchy, State, and Utopia* (New York: Basic Books, 1974).

Nyman, John A., "The Economics of Moral Hazard Revisited," *Journal of Health Economics 18* (1999): 811–824.

Nyman, John A., "Testing for Excess Demand in Nursing Home Markets," *Medical Care 31* (1993): 680–693.

Nyman, John A., and Roland Maude-Griffin, "The Welfare Economics of Moral Hazard," *International Journal of Health Care Finance and Economics 1* (2001): 23–42.

Nyman, John A., et al., "Changing Physician Behavior: Does Medical Review of Part B Medicare Claims Make a Difference?" *Inquiry 27* (1990): 127–137.

O'Halloran, Patrick L., and David J. Bashaw, "U.S. Physician Board Certification and Labor Market Returns," *International Journal of Manpower 27* (2006): 624–646.

Ohsfeldt, Robert L., Raymond G. Boyle, and Eli I. Capilouto, "Tobacco Taxes, Smoking Restrictions, and Tobacco Use," *NBER* Working Paper 6486, 1998.

Okun, Arthur, *Efficiency and Equity: The Big Tradeoff* (Washington, DC: Brookings, 1975).

Okunade, Ablert A., and Chutima Suraratdecha, "Cost Efficiency, Factor Interchange, and Technical Progress in U.S. Specialized Hospital Pharmacies," *Health Economics 7* (1998): 363–371.

Oldfield, Gary, and Susan Eaton, *Dismantling Regulation* (New York: New Press, 1996).

Olesen, Ole B., and Niels C. Petersen, "The Use of Data Envelopment Analysis with Probablistic Assurance Regions for Measuring Hospital Efficiency," *Journal of Productivity Analysis 17* (2002): 83–109.

Olivella, Pau, and Marcos Vera-Hernández, "Competition among Differentiated Health Plans under Adverse Selection," *Journal of Health Economics 26* (2007): 233–250.

Olkin, Ingram, and Jerome Sacks, "Cross-Disciplinary Research in the Statistical Sciences," Hayward, CA: Institute of Mathematical Statistics, September 1988.

Olsen, Jan Abel, "But Health Can Still Be a Necessity. . . . ," *Journal of Health Economics 12* (1993): 187–191.

Olsen, Jan Abel, "Theories of Justice and Their Implications for Priority Setting in Health Care," *Journal of Health Economics 16* (1997): 625–639.

Olsen, Jan Abel, and Richard D. Smith, "Theory versus Practice: A Review of 'Willingness-to-Pay' in Health and Health Care," *Health Economics 10* (2001): 39–52.

Olson, Mary K., "The Risk We Bear: The Effects of Review Speed and Industry User Fees on New Drug Safety," *Journal of Health Economics 27* (2008): 175–200.

O'Neill, June E., and Dave M. O'Neill, "Health Status, Health Care and Inequality: Canada vs. the U.S.," *Forum for Health Economics & Policy 10* (2008) (Frontiers in Health Policy Research): Article 3. www.bepress.com/fhep/10/1/3, accessed November 2016.

Oreopoulus, Philip, "Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter," *American Economic Review 96* (2006): 125–175.

Organization for Economic Cooperation and Development (OECD), *OECD Health Data*, various years, various tables, 75775 Paris Cedex 16, France.

Organization for Economic Cooperation and Development (OECD), "Health Care Systems: Getting More Value for Money," *OECD Economics Department Policy Notes, No. 2*, 2010.

Orzechowski, William, and Robert C. Walker, *The Tax Burden on Tobacco, Historical Compilation*, Vol. 49, 2014. www.taxadmin.org/assets/docs/Tobacco/papers/tax_burden_2014.pdf, accessed September 2016.

Owen, Gary M., "Seizing the Opportunity," *American Health & Drug Benefits* (April 2008): 52–55.

Parente, Stephen, and Roger Feldman, "Consumer Response to a National Marketplace for Individual Insurance," presented at the American Enterprise Institute, Washington, DC, July 31, 2008.

Paris, Valérie, Marion Devaux, and Lihan Wei, "Health Systems Institutional Characteristics: A Survey of 29 OECD Countries," *OECD Health Working Paper No. 50*, OECD Publishing, 2010.

Parker-Pope, Tara, "Antismoking Sentiment Flares in Europe's Smoke-Filled Cafes," *Wall Street Journal*, August 28, 1995, sec. B1.

Parkin, David, Alistair McGuire, and Brian Yule, "Aggregate Health Care Expenditures and National Income: Is Health Care a Luxury Good?" *Journal of Health Economics 6* (1987): 109–128.

Paul, Chris, "Physician Licensure and the Quality of Medical Care," *Atlantic Economic Journal 12* (1984): 18–30.

Pauly, Mark V., "The Economics of Moral Hazard: Comment," *American Economic Review 58* (1968): 531–537.

Pauly, Mark V., "Editorial: A Re-Examination of the Meaning and Importance of Supplier-Induced Demand," *Journal of Health Economics 13* (1994): 369–372.

Pauly, Mark V., *Health Benefits at Work: An Economic and Political Analysis of Employment-Based Health Insurance* (Ann Arbor, MI: The University of Michigan Press, 1997).

Pauly, Mark V., "Is Medical Care Different?" in *Competition in the Health Care Sector*, Warren Greenberg (ed.) (Germantown, MD: Aspen Systems, 1978).

Pauly, Mark V., "Is Medical Care Different? Old Questions, New Answers," *Journal of Health Politics, Policy and Law 13* (1988): 227–237.

Pauly, Mark V., *Medical Care at Public Expense* (New York: Praeger, 1971).

Pauly, Mark V., and Michael Redisch, "The Not-for-Profit Hospital as a Physician's Cooperative," *American Economic Review 63* (1973): 87–100.

Pauly, Mark V., and Uwe E. Reinhardt, "The Rise and Fall of Health Care Reform," in *Looking Back, Looking Forward, Staying Power Issues in Health Care Reform* (Washington, DC: National Academies Press, 1996).

Pauly, Mark V., and Mark A. Satterthwaite, "The Pricing of Primary Care Physicians' Services: A Test of the Role of Consumer Information," *Bell Journal of Economics 12* (1981): 488–506.

Payne, Abigail A., "Does the Government Crowd Out Private Donations? New Evidence from a Sample of Non-Profit Firms," *Journal of Public Economics 69* (1998): 323–345.

Pearson, Steven D., et al., "The Impact of Pay-for-Performance on Health Care Quality in Massachusetts, 2001–2003," *Health Affairs 27* (2008): 1167–1176.

Peckham, Carol, "Medscape Physician Compensation Report, 2015," April 21, 2015: available at www.medscape.com/features/slideshow/compensation/2015/public/overview, accessed November 2016.

Peltzman, Sam, *Regulation of Pharmaceutical Innovation: The 1962 Amendments*, American Enterprise Institute, Washington, DC, 1974.

Peltzman, Sam, "Toward a More General Theory of Regulation," *Journal of Law and Economics 19* (1976): 211–248.

Pernick, Martin S., "The Cultural Ecology of Health: Science, Values, and Society in the History of American Medicine," manuscript of book in progress, 1991.

Peterson, Mark A. (ed.), "Special Issue: Kenneth Arrow and the Changing Economics of Health Care," *Journal of Health Politics, Policy and Law 26* (October 2001).

Pharmaceutical Research and Manufacturers of America (PhRMA), *The Facts about Pharmaceutical Marketing and Promotion*, Washington, DC, 2008 (available at www.phrma.org, accessed November 2016).

Pharmaceutical Research and Manufacturers of America (PhRMA), *New Drug Approvals in 2009*, Washington, DC, 2010 (available at www.phrma.org, accessed November 2016).

Pharmaceutical Research and Manufacturers of America (PhRMA), *Pharmaceutical Industry Profile 2015*, Washington, DC, 2015 (available at www.pharma.org, accessed November 2016).

Phelps, Charles E., "Good Technologies Gone Bad: How and Why the Cost-Effectiveness of a Medical Intervention Changes for Different Populations," *Medical Decision Making 17* (1997): 107–117.

Phelps, Charles E., "Induced Demand—Can We Ever Know Its Extent?" *Journal of Health Economics 5* (1986): 355–365.

Phelps, Charles E., "Information Diffusion and Best Practice," in *Handbook of Health Economics*, Volume I, Anthony J. Culyer and Joseph P. Newhouse (eds.) (Amsterdam: Elsevier Science B.V., 2000).

Phelps, Charles E., "Welfare Loss from Variations: Further Considerations," *Journal of Health Economics 14* (1995): 253–260.

Philipson, Tomas J., Essy Mozaffari, and John R. Maclean, "Pharmacy Cost Sharing, Antiplatelet Therapy Utilization, and Health Outcomes for Patients with Acute Coronary Syndrome," *American Journal of Managed Care 16* (2010): 290–297.

Philipson, Tomas, et al., "Cost-Benefit Analysis of the FDA: The Case of the Prescription Drug User Fee Acts," *Journal of Public Economics 92* (2008): 1306–1325.

Phillips, Kathryn A., David Schleifer, and Carolin Hagelskamp, "Most Americans Do Not Believe That There is an Association between Health Care Prices and Quality of Care," *Health Affairs 35* (2016): 647–653.

Photius.com, "GDP – Purchasing Power Parity 2008 Country Ranks," www.photius.com/rankings/economy/gdp_purchasing_power_parity_2008_0.html, accessed November 2016.

Picone, Gabriel, Shin-Yi Chou, and Frank Sloan, "Are For-Profit Hospital Conversions Harmful to Patients and to Medicare?" *RAND Journal of Economics 33* (2002): 507–523.

Picone, Gabriel A., Martin Uribe, and R. Mark Wilson, "The Effect of Uncertainty on the Demand for Medical Care, Health Capital and Wealth," *Journal of Health Economics 17* (1998): 171–185.

Pinkovsky, Maxim, and Xavier Sala-i-Martin, "Parametric Estimation of the World Distribution on Income," Working Paper 15433, National Bureau of Economic Research, 2009.

Plotzke, Michael Robert, and Courtemanche, Charles, "Does Procedure Profitability Impact Whether an Outpatient Surgery Is Performed at an Ambulatory Surgery Center or Hospital?" *Health Economics 20* (2011): 817–830.

Pollard, Michael R., and Robert F. Leibenluft, *Antitrust and the Health Professions*, Office of Policy and Planning, Federal Trade Commission, 1981.

Pollitz, Karen, and Mathew Rae, "Workplace Wellness Programs Characteristics and Requirements," Kaiser Family Foundation Issue Brief, January 2016.

Polsky, Daniel, and Sean Nicholson, "Why Are Managed Care Plans Less Expensive: Risk Selection, Utilization, or Reimbursement?" *Journal of Risk and Insurance 71* (2004): 21–40.

Polsky, Daniel, et al., "Malpractice Premiums and the Supply of Obstetricians," *Inquiry 47* (2010): 48–61.

Porell, Frank, et al., "A Longitudinal Analysis of Nursing Home Outcomes," *Health Services Research 33* (1998): 835–865.

Porter, Eduardo, "Do New Drugs Always Have to Cost So Much?" *New York Times*, November 14, 2004; Section 3, p. 5.

Powell, Lisa M., and Frank T. Chaloupka, "Food Prices and Obesity: Evidence and Implications for Taxes and Subsidies," *The Milbank Quarterly 87* (2009): 229–257.

Pozen, Alexis, and David M. Cutler, "Medical Spending Differences in the United States and Canada: The Role of Prices, Procedures, and Administrative Expenses," *Inquiry 47* (2010): 124–134.

President of the United States, *The President's Comprehensive Health Reform Program*, February 6, 1992.

Preya, Colin, and George Pink, "Scale and Scope Efficiencies through Hospital Consolidations," *Journal of Health Economics 25* (2006): 1049–1068.

Pritchett, Lant, and Lawrence H. Summers, "Wealthier Is Healthier," *Journal of Human Resources 31* (1996): 841–868.

Putnam, Robert D., *Bowling Alone* (New York: Simon & Schuster, 2000).

Putnam, Robert, "E Pluribus Unum: Diversity and Community in the Twenty-First Century," *Scandinavian Political Studies* 30 (2007): 137–174.

Putnam, Robert, *Our Kids: The American Dream in Crisis* (New York: Simon & Schuster, 2015).

Qin, Paige, and Michael Chernew, "Compensating Wage Differentials and the Impact of Health Insurance in the Public Sector on Wages and Hours," *Journal of Health Economics 38* (2014) 77–87.

Quast, Troy, David Sappington, and Elizabeth Shenkman, "Does the Quality of Care in Medicaid MCOs Vary with the Form of Physician Compensation?" *Health Economics 17* (2008): 545–550.

Raffel, Marshall W., Norma K. Raffel, and Camille K. Barsukiewicz, *The U.S. Health System: Origins and Functions*, 5th ed. (Albany, New York: Delmar, 2002).

Rashad, Inas, and Eric Sarpong, "Employer-Provided Health Insurance and the Incidence of Job Lock: A Literature Review and Empirical Test," *Expert Review of Pharmacoeconomics & Outcomes Research 8* (2008): 583–591.

Rawls, John, *A Theory of Justice* (Cambridge, MA: Belnap Press of the Harvard University Press, 1971).

Reardon, Sean, "The Widening Academic Gap between the Rich and the Poor," in *Wither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, (eds.) Greg J. Duncan and Richard Murnane, (New York: Russell Sage Foundation, 2011).

Register, Charles A., and Edward R. Bruning, "Profit Incentives and Technical Efficiency in the Production of Hospital Care," *Southern Economic Journal 53* (1987): 899–914.

Reinhardt, Uwe E., "Editorial: The Theory of Physician-Induced Demand: Reflections after a Decade," *Journal of Health Economics* (1985).

Reinhardt, Uwe E., "The Pricing of U.S. Hospital Services: Chaos behind a Veil of Secrecy," *Health Affairs 25* (2006): 57–69.

Reinhardt, Uwe E., "A Production Function for Physician Services," *Review of Economics and Statistics 54* (1972): 55–65.

Relman, Arnold S., and Marcia Angell, "America's Other Drug Problem," *The New Republic* (December 16, 2002): 27–41.

Renn, Steven C., et al., "The Effects of Ownership and System Affiliation on the Economic Performance of Hospitals," *Inquiry 22* (1985): 219–236.

Research Data Assistance Center (ResDAC), "Calculating Hospital Specific DRG Adjusted Payments," University of Minnesota, www.resdac.umn.edu/Tools/TBs/TN-004-2.pdf, accessed July 31, 2008.

Reyes, Jessica Wolpaw, "Do Female Physicians Capture Their Scarcity Value: The Case of OB/GYNs," *NBER* Working Paper 12528, 2006.

Rhee, Sang-O et al., "USMGs versus FMGs: Are There Performance Differences in the Ambulatory Setting?" *Medical Care 24* (1986): 248–258.

Rice, Dorothy, *Estimating the Cost of Illness*, U.S. Public Health Service Publication No. 947–6 (Washington, DC: U.S. Government Printing Office, May 1966).

Rice, Thomas, *The Economics of Health Reconsidered* (Chicago: Health Administration Press, 1998).

Rice, Thomas H., and Roberta J. Labelle, "Do Physicians Induce Demand for Medical Service?" *Journal of Health Politics, Policy and Law 14* (1989): 587–600.

Ried, Walter, "QALYs versus HYEs—What's Right and What's Wrong: A Review of the Controversy," *Journal of Health Economics 17* (1998): 607–625.

Rizzo, John A., "Are HMOs Bad for Health Maintenance?" *Health Economics 13* (2005): 1117–1131.

Rizzo, John A., "Physician Advertising Revisited," *Medical Care 26* (1988): 1238–1244.

Rizzo, John A., and David Blumenthal, "Is the Target Income Hypothesis an Economic Heresy?" *Medical Care Research and Review 53* (1996): 243–266.

Rizzo, John A., and Richard J. Zeckhauser, "Advertising and Entry: The Case of Physician Services," *Journal of Political Economy 98* (1990): 476–500.

Rizzo, John A., and Richard J. Zeckhauser, "Advertising and the Price, Quantity, and Quality of Primary Care Physician Services," *Journal of Human Resources 27* (1992): 381–421.

Rizzo, John, and Richard Zeckhauser, "Pushing Income to Reference Points: Why Do Male Doctors Earn More?" *Journal of Economic Behavior and Organization 63* (2007): 514–536.

Rizzo, John A., and Richard J. Zeckhauser, "Reference Incomes, Loss Aversion and Physician Behavior," *Review of Economics and Statistics 85* (2003): 909–922.

Roberts, Brandon, and Irving Hoch, "Malpractice Litigation and Medical Costs in the United States," *Health Economics 18* (2009): 1394–1419.

Roberts, Mark J., and Larry Samuelson, "An Empirical Analysis of Dynamic, Nonprice Competition in an Oligopolistic Industry," *RAND Journal of Economics 19* (1988): 200–220.

Robinson, James C., "Hospitals Respond to Medicare Payment Shortfalls by Both Shifting Costs and Cutting Them, Based on Market Concentration," *Health Affairs 30* (2011): 1265–1271.

Rocco, Lorenzo, and Béatrice d'Hombres, "Social Capital and Smoking," in *The Economics of Social Capital and Health*, S. Folland and L. Rocco (eds.), 177–202 (Singapore: World Scientific, 2014).

Rocco, Lorenzo, E. Fumagalli, and M. Suhrcke, "From Social Capital to Health – and Back," *Health Economics 23* (2013): 586–605.

Rochaix, Lise, "Information Asymmetry and Search in the Market for Physicians' Services," *Journal of Health Economics 8* (1989): 53–84.

Roehrig, Charles S., "Health Sector Trend Report, June 2016," Ann Arbor: Altarum Institute, www.altarum.org/cshs, accessed July 18, 2016.

Ronconi, Lucas, Timothy T. Brown, and Richard M. Scheffler, "Social Capital and Self Related Health in Argentina," *Health Economics 21* (2010): 201–208.

Rosenthal, Jaime A., Xin Lu, and Peter Cram, "Availability of Consumer Prices from U.S. Hospitals for a Common Surgical Procedure," *JAMA: Internal Medicine 173* (2013): 427–432.

Rosenthal, Meredith B., et al., "Early Experience with Pay-for-Performance: From Concept to Practice," *Journal of the American Medical Association 294* (2005): 1788–1793.

Rosenthal, Meredith B., et al., "Pay for Performance in Commercial HMOs," *New England Journal of Medicine 355* (2006): 1895–1902.

Rosenzweig, Mark R., and T. Paul Schultz, "Estimating a Household Production Function: Heterogeneity, The Demand for Health Inputs, and Their Effects on Birth Weight," *Journal of Political Economy 91* (1983): 723–746.

Rosenzweig, Mark R., and Kenneth I. Wolpin, "Sisters, Siblings, and Mothers: The Effect of Teen-Age Childbearing on Birth Outcomes in a Dynamic Family Context," *Econometrica 63* (1995): 303–326.

Ruffin, Roy J., and Paul R. Gregory, *Principles of Economics*, 7th ed. (New York: Addison Wesley, 2001).

Ruger, Jennifer P., et al., "Costs of Addressing Heroin Addiction in Malaysia and 32 Comparable Countries Worldwide," *Health Services Research 47* (2012): 865–887.

Ruggeri, Kai, and Ellen Nolte, *Pharmaceutical Pricing: The Use of External Pricing Reference*, RAND Corporation, 2013.

Saffer, Henry, "Alcohol Advertising and Motor Vehicle Fatalities," *Review of Economics and Statistics 79* (1997): 431–442.

Saffer, Henry, "Alcohol Advertising Bans and Alcohol Abuse: An International Perspective," *Journal of Health Economics 10* (1991): 65–79.

Saffer, Henry, "Alcohol Consumption and Alcohol Advertising Bans," *NBER* Working Paper 7758, 2000.

Saffer, Henry, and Frank Chaloupka, "Demographic Differentials in the Demand for Alcohol and Illicit Drugs," *NBER* Working Paper 6432, 1998.

Saffer, Henry, and Dhaval Dave, "Alcohol Advertising of Alcohol Consumption by Adolescents," *NBER* Working Paper 9676, 2003.

Saffer, Henry, and Dhaval Dave, "Mental Illness and the Demand for Alcohol, Cocaine, and Cigarettes," *Economic Inquiry 43* (2005): 229–246.

Safriet, Barbara J., "Impediments to Progress in Health Care Workforce Policy: License and Practice Laws," *Inquiry 31* (1994): 310–317.

Santerre, Rexford, "The Welfare Loss from Hospital Cost-Shifting Behavior: A Partial Equilibrium Analysis," *Health Economics 14* (2005): 621–626.

Santerre, Rexford E., and John A. Vernon, "The Consumer Welfare Implications of the Hospital Ownership Mix in the U.S.: An Exploratory Study," *Health Economics 15* (2006): 1187–1199.

Sanz de Galdeano, Anna, "Job-Lock and Public Policy: Evidence from Clinton's Second Mandate," *Industrial and Labor Relations Review 59* (2006): 430–437.

Sapolsky, Robert M., *Why Zebras Don't Get Ulcers: An Updated Guide to Stress Related Diseases and Coping* (New York: W.H. Freeman and Company, 1998).

Sasser, Alicia C., "Gender Differences in Physician Pay: Tradeoffs between Career and Family," *Journal of Human Resources 60* (2005): 477–504.

Satterthwaite, Mark A., "Consumer Information, Equilibrium Industry Price, and the Number of Sellers," *Bell Journal of Economics 10* (1979): 483–502.

Satvanath, Shaker, Nico Voigtlaender, and Hans-Joachim Voth, "Bowling for Fascism: Social Capital and the Rise of the Nazi Party," *NBER* Working Paper 19201, 2013.

Scanlon, Dennis P., et al., "Competition and Health Plan Performance: Evidence from Health Maintenance Organization Insurance Market," *Medical Care 43* (2005): 338–346.

Scanlon, Dennis P., et al., "The Impact of Health Plan Report Cards on Managed Care Enrollment," *Journal of Health Economics 21* (2002): 19–41.

Scanlon, William J., "A Theory of the Nursing Home Market," *Inquiry 17* (1980): 25–41.

Schaefer, Mark E., "Demand versus Need for Medical Services in a General Cost Benefit Setting," *American Journal of Public Health 65* (1975): 293–295.

Schelling, Thomas, "Ergonomics: The Art of Self Management," *American Economic Review 68* (1978): 290–294.

Scherer, F. Michael, *Innovation and Growth: Schumpeterian Perspectives* (Cambridge, MA: MIT, 1984).

Schmidt, Peter, and Robin Sickles, "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics 2* (1984): 367–374.

Schneeweiss, Sebastian, et al., "The Effect of Medicare Part D Coverage on Drug Use and Cost Sharing among Seniors without Prior Drug Benefits," *Health Affairs* 28 (2009): w305–w316.

Schnelle, John F., et al., "Relationship of Nursing Home Staffing to Quality of Care," *Health Services Research 39* (2004): 225–250.

Schoen, Cathy, "The Affordable Care Act and the U.S. Economy: A Five-Year Perspective," New York: Commonwealth Foundation, February 2016, www.commonwealthfund.org/publications/fund-reports/2016/feb/aca-economy-five-year-perspective, accessed July 18, 2016.

Schoen, Cathy, et al., "How Health Insurance Design Affects Access to Care and Costs by Income in Eleven Countries," *Health Affairs 29* (2010): 2323–2334.

Schwartz, Joel, and Douglas W. Dockery, "Increased Mortality in Philadelphia Associated with Daily Air Pollution Concentrations," *American Review of Respiratory Diseases 145* (1992): 600–604.

Schwartz, Michael, et al., "Small Area Variations in Hospitalization Rates: How Much You See Depends on How You Look," *Medical Care 32* (1994): 189–201.

Schwartz, William B., "In the Pipeline: A Wave of Valuable Medical Technology," *Health Affairs 13* (1994): 70–79.

Schwartz, William B., and Daniel N. Mendelson, "Eliminating Waste and Inefficiency Can Do Little to Contain Costs," *Health Affairs 13* (1994): 224–238.

Schweitzer, Stuart O., *Pharmaceutical Economics and Policy* (New York: Oxford University Press, 1996).

Seldon, Barry J., and Khowsrow Doroodian, "A Simultaneous Model of Cigarette Advertising: Effects on Demand and Industry Response to Public Policy," *Review of Economics and Statistics 71* (1989): 673–677.

Seldon, Barry J., Chulho Jung, and Roberto J. Cavazos, "Market Power among Physicians in the U.S., 1983–1991," *Quarterly Review of Economics and Finance 38* (1998): 799–824.

Sen, Amartya, "Equality of What?" in *The Tanner Lectures on Human Values*, S. McMurrin (ed.) (Cambridge: Cambridge University Press, 1980).

Sen, Anindya, and Tony S. Wirjanto, "Estimating the Impacts of Taxes on the Initiation and Persistence of Youth Smoking: Empirical Evidence from Ontario, Canada," *Health Economics 19* (2010): 1264–1280.

Sen, Bisakha, and Christine M. Campbell, "Alcohol Prevalence, Alcohol Policies, and Child Fatal Injury Rates from Motor Vehicle Crashes," *Contemporary Economic Policy 28* (2010): 392–405.

Sendi, Pedram, J. Al Maiwenn, and Heinz Zimmerman, "A Risk-adjusted Approach to Comparing the Return on Investment in Health Care Programs," *International Journal of Health Care Finance and Economics 4* (2004): 199–210.

Sharma, Rajiv, Arnab Mitra, and Miron Stano, "Insurance, Race/Ethnicity and Sex in the Search for a New Doctor: ACA-Baseline Estimates from an Audit Study," *Economics Letters 37* (2015): 150–153.

Sharma, Rajiv, et al., "Access to Primary Care Physicians: Results from a National Audit Study," presented at the 91st Annual Conference of the Western Economic Association International, July 2, 2016.

Sheingold, Steven, Nguyen Nguyen, and Andre Chappel, "Competition and Choice in Health Insurance Marketplaces, 2014–2015: Impact on Premiums," ASPE Issue Brief, Department of Health and Human Services, July 27, 2015.

Shen, Yu-chu, "The Effect of Financial Pressure on the Quality of Care in Hospitals," *Journal of Health Economics 22* (2003): 243–269.

Sherman, H. David, "Hospital Efficiency Measurement and Evaluation," *Medical Care 22* (1984): 922–935.

Sheu, Mei-ling, et al., "The Effect of a Major Cigarette Price Change on Smoking Behavior in California: A Zero-Inflated Negative Binomial Model," *Health Economics 13* (2004): 781–791.

Shleifer, Andrei, "A Theory of Yardstick Competition," *RAND Journal of Economics 16* (1985): 319–327.

Shortell, Stephen M., Robin R. Gillies, and David A. Anderson, "The New World of Managed Care: Creating Organized Delivery Systems," *Health Affairs 13* (1994): 46–64.

Sickles, Robin C., and Abdo Yazbeck, "On the Dynamics of Demand for Leisure and the Production of Health," *Journal of Business and Economic Statistics 16* (1998): 187–197.

Siddique, Haroon, "NHS Hospital Waiting Time Figures Worst in Seven Years," *The Guardian*, April 9, 2015, www.theguardian.com/society/2015/apr/09/nhs-hospital-waiting-time-figures, last accessed June 3, 2016.

Silverman, Elaine, and Jonathan Skinner, "Are For Profit Hospitals Really Different? Medicare Upcoding and Market Structure," *NBER* Working Paper 8133, 2001.

Silverman, Elaine, and Jonathan Skinner, "Medicare Upcoding and Hospital Ownership," *Journal of Health Economics 23* (2004): 369–389.

Simon, Carol J., David Dranove, and William White, "The Effect of Managed Care on the Incomes of Primary and Specialty Physicians," *Health Services Research 33* (1998): 549–569.

Simon, Kosali I., "Nation's Big Uninsurance Problem Revolves around Small Firms," Economic Research Initiative on the Uninsured, eriu.sph.umich.edu/forthemedia/interviews_simon.html, accessed September 1, 2008.

Sirven, N., and Delbrand, T., "Social Capital and Health of Older Europeans: Causal Pathways and Health Inequalities," *Social Science & Medicine* 75 (2012): 1288–1295.

Skinner, Jonathan, "Causes and Consequences of Geographic Variation in Health Care," in *Handbook of Health Economics*, Volume 2, Mark V. Pauly, Thomas McGuire, and Pedro P. Barros (eds.) (Amsterdam: Elsevier Science B.V., 2012).

Skinner, Jonathan, "What Do Stochastic Cost Functions Tell Us about Inefficiency?" *Journal of Health Economics 13* (1994): 323–328.

Sloan, Frank A., "Property Rights in the Hospital Industry," in *Health Care in America*, H. E. Frech III (ed.) (San Francisco: Pacific Research Institute for Public Policy, 1988).

Sloan, Frank A., "Rate Regulation as a Strategy for Hospital Cost Control: Evidence from the Last Decade," *Milbank Memorial Fund Quarterly/Health and Society 61* (1983): 195–221.

Sloan, Frank A., and Edmund R. Becker, "Internal Organization of Hospitals and Hospital Costs," *Inquiry 28* (1981): 224–239.

Sloan, Frank A., Michael A. Morrisey, and Joseph Valvona, "Effects of the Medicare Prospective Payment System on Hospital Cost Containment: An Early Appraisal," *Milbank Quarterly 66* (1988a): 191–219.

Sloan, Frank A., Michael A. Morrisey, and Joseph Valvona, "Medicare Prospective Payment and the Use of Medical Technologies in Hospitals," *Medical Care 26* (1988b): 837–853.

Sloan, Frank A., et al., "The Demise of Hospital Philanthropy," *Economic Inquiry 28* (1990): 725–743.

Sloan, Frank A., et al., "Does Where You Are Admitted Make a Difference? An Analysis of Medicare Data," *NBER* Working Paper 6896, 1999.

Smedley, Brian D., Adrienne Y. Stith, and Alan R. Nelson (eds.), *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Institute of Medicine, National Academy Press, Washington, DC, 2002.

Smith, Adam, *An Inquiry into the Nature and Causes of the Wealth of Nations* (Chicago: University of Chicago Press, 1976).

Smith, Francis B., *The Retreat of Tuberculosis 1850–1950* (London: Croom Helm, 1988).

Smith, James P., "Socioeconomic Status and Health," *American Economic Review 88* (1998): 192–196.

Smith, Michael J., Susan S. Ellenberg, Louis M. Bell, and David M. Rubin, "Media Coverage of the Measles-Mumps-Rubella Vaccine and Autism Controversy and Its Relationship to MMR Immunization Rates in the United States," *Pediatrics 121* (2008): e836–e843.

Sommers, Benjamin D., et al., "Changes in Self-Reported Insurance Coverage, Access to Care, and Health under the Affordable Care Act," *Journal of the American Medical Association 314* (2015): 366–374.

Sommers, Benjamin D., et al., "Changes in Utilization and Health among Low-Income Adults after Medicaid Expansion or Expanded Private Insurance," *JAMA Internal Medicine*: Published Online August 8, 2016: doi:10.1001/jamaintermed.2016.4419.

Sonnenfeld, Sally T., et al., "Projections of National Health Expenditures through the Year 2000," *Health Care Financing Review 13* (1991): 1–27.

Spence, A. Michael, "Monopoly, Quality, and Regulation," *Bell Journal of Economics 6* (1973): 417–429.

Stano, Miron, "An Analysis of the Evidence on Competition in the Physician Services Markets," *Journal of Health Economics 4* (1985): 197–211.

Stano, Miron, "A Further Analysis of the Physician Inducement Controversy," *Journal of Health Economics 6* (1987): 227–238.
Stano, Miron, "Individual Health Accounts: An Alternative Health Care Financing Approach," *Health Care Financing Review 3* (1981): 117–125.
Stano, Miron, and Sherman Folland, "Variations in the Use of Physician Services by Medicare Beneficiaries," *Health Care Financing Review 9* (1988): 51–58.
Stano, Miron, et al., "The Effects of Physician Availability on Fees and the Demand for Doctors' Services," *Atlantic Economic Journal 13* (1985): 51–60.
Starr, Paul, "Health Care for the Poor: The Past Twenty Years," in *Fighting Poverty: What Works and What Doesn't*, Sheldon Danziger and Daniel Weinberg (eds.) (Cambridge: Harvard, 1986).
Steinberg, Richard, "Charitable Giving as a Mixed Public/Private Good: Implications for Tax Policy," *Public Finance Quarterly 14* (1986): 415–431.
Steinberg, Richard, "Voluntary Donations and Public Expenditures in a Federalist System," *American Economic Review 77* (1987): 24–36.
Steinmann, Lukas, and Peter Zweifel, "On the [In]efficiency of Swiss Hospitals-English Summary," *Ifo Studien 46* (2000): 197–217.
Stern, Nicholas, *The Economics of Climate Change: The Stern Review* (Cambridge: Cambridge University Press, 2007).
Stigler, George J., "The Economics of Information," *Journal of Political Economy 69* (1961): 213–225.
Stigler, George J., "The Theory of Economic Regulation," *Bell Journal of Economics 2* (1971): 3–21.
Stiglitz, Joseph E., *Economics of the Public Sector*, 2nd ed. (New York: Norton, 1988).
Stiglitz, Joseph E., *The Price of Inequality: How Today's Divided Society Endangers our Future* (New York: Norton, 2012).
Street, Andrew, "How Much Confidence Can We Have in Hospital Efficiency Estimates?" *Health Economics 12* (2003): 896–907.
Succi, Melissa J., Shoou-Yih Lee, and Jeffrey A. Alexander, "Effects of Market Position and Competition on Rural Hospital Closures," *Health Services Research 31* (1997): 679–699.
Surgeon General of the United States, *Reducing the Health Consequences of Smoking: 25 Years of Progress*, *Executive Summary*. U.S. Department of Health and Human Services, Public Health Service, Office on Smoking and Health, Rockville, MD, 1989.
Svorny, Shirley, "Medical Licensing: An Obstacle to Affordable, Quality Care," *Policy Analysis No. 621*, Cato Institute, September 17, 2008.
Svorny, Shirley, "Should We Reconsider Licensing Physicians?" *Contemporary Policy Issues 10* (1992): 31–38.
Tak Hyo-Jung, Gregory W. Ruhnke, and David O. Meltzer, "Association of Patient Preferences for Participation in Decision Making with Length of Stay and Costs among Hospitalized Patients," *JAMA Internal Medicine 173* (2013): 1195–1203.
Takakura, M., "Does Social Trust at School Affect Students' Smoking and Drinking Behavior in Japan?" *Social Science & Medicine 72* (2011): 299–306.
Tampubolon, G, S.V. Subramanian, and I. Kawachi, "Neighborhood Social Capital and Individual Self-Related Health in Wales," *Health Economics 22* (2011): 14–21.
Taubman, Sarah L., et al., "Medicaid Increases Emergency Department Use: Evidence from Oregon's Health Insurance Experiment," *Science 343* (2014): 263–269.
Tauras, John A., and Frank J. Chaloupka, "Price, Clean Indoor Air, and Cigarette Smoking: Evidence from the Longitudinal Data for Young Adults," *NBER* Working Paper 6937, 1999.

Tay, Abigail, "Assessing Competition in Hospital Care Markets: The Importance of Accounting for Quality Difference," *RAND Journal of Economics 34* (2003): 786–814.

Temin, Peter, "An Economic History of American Hospitals," in *Health Care in America*, H. E. Frech III (ed.) (San Francisco: Pacific Research Institute for Public Policy, 1988).

Teplensky, Jill D., et al., "Hospital Adoption of Medical Technology: An Empirical Test of Alternative Models," *Health Services Research 30* (1995): 437–466.

Thaler, Richard, and Sherwin Rosen, "The Value of Saving a Life: Evidence from the Labor Market," in *Household Production and Consumption*, Nestor E. Terleckyi (ed.) (New York: National Bureau of Economic Research, 1975): 265–297.

Thornton, James, "Estimating a Health Production Function for the U.S.," *Applied Econometrics 34* (2002): 59–62.

Tilford, John M., et al., "Response to Health Insurance by Previously Uninsured Rural Children," *Health Services Research 34* (1999): 761–775.

Tinkler, Sarah, et al., "Offers of Appointments with Nurse Practitioners if a Requested Physician is Unavailable," *Journal of the American Association of Nurse Practitioners* 2016.

Torrance, George W., "Preferences for Health Outcomes and Cost-Utility Analysis," *American Journal of Managed Care 3* (1997 Supplement): S8–S20.

Toxic Substances Board, *Health or Tobacco: An End to Tobacco Advertising and Promotion*, Department of Health, Wellington, New Zealand, May 1989.

Tremblay, Carol Horton, and Victor J. Tremblay, "The Impact of Cigarette Advertising on Consumer Surplus, Profit, and Social Welfare," *Contemporary Economic Policy 13* (1995): 113–124.

Troyer, Jennifer L., "Examining Differences in Death Rates for Medicaid and Non-Medicaid Nursing Home Residents," *Medical Care 42* (2004): 985–991.

Trude, Sally, Melanie Au, and Jon B. Christianson, "Health Plan Pay-for-Performance Strategies," *American Journal of Managed Care 12* (2006): 537–542.

Tsuchiya, Aki, "QALYs and Ageism: Philosophical Theories and Age Weighting," *Health Economics 9* (2000): 57–68.

Tumlinson, Anne, et al., "Choosing a Health Plan: What Information Will Consumers Use?" *Health Affairs 16* (1997): 229–238.

Tye, Joe B., Kenneth E. Warner, and Stanton A. Glantz, "Tobacco Advertising and Consumption: Evidence of a Causal Relationship," *Journal of Public Health Policy 8* (1987): 492–508.

U.S. Department of Commerce, *Statistical Abstract*. Washington, DC, various years.

U.S. Department of Health and Human Services, *Healthy People 2010: Understanding and Improving Health*, 2nd edn., US Government Printing Office, 2000.

U.S. Department of Health and Human Services, *Understanding Estimates of the Uninsured: Putting the Differences in Context*, http://aspe.os.dhhs.gov/health/reports/hiestimates.htm, accessed March 17, 2003.

U.S. Department of Health and Human Services, Medicare Payment Advisory Commission, Medicare and the Health Care Delivery System, Washington, DC, June 2015.

U.S. Department of Health and Human Services, Public Health Service, *Seventh Report to the President and Congress on the Status of Health Personnel in the United States*. DHHS Publication No. HRS-P-OD-90-1, March 1990.

U.S. Department of Health and Human Services, Public Health Service. *Health United States*, various years.

Valdez, R. Burciaga et al., "Consequences of Cost Sharing for Children's Health," *Pediatrics 75* (1985): 952–961.

Valdmanis, Vivian G., "Ownership and Technical Efficiency of Hospitals," *Medical Care 28* (1990): 552–561.

Van der Wees, Philip, A.M. Zaslavsky, and J.Z. Ayanian, "Improvements in Health Status after Massachusetts Health Care Reform," *Milbank Quarterly 91* (2013): 663–689.
Van Doorslaer, Eddy, Xander Koolman, and Andrew M. Jones, "Explaining Income-Related Inequalities in Doctor Utilization in Europe," *Health Economics 13* (2004): 629–647.
Van Houtven, Courtney, and Edward C. Norton, "Informal Care and Health Use of Older Adults," *Journal of Health Economics 23* (2004): 1159–1180.
Van Houtven, Courtney H., and Edward C. Norton, "Informal Care and Medical Expenditures: Testing for Heterogeneous Treatment Effects," *Journal of Health Economics 27* (2008): 134–156.
Van Lear, William, and Lynette Fowler, "Efficiency and Service in the Group Home Industry," *Journal of Economic Issues 21* (1997): 1039–1050.
Varkevisser, Marco, Stéphanie A. van der Geest, and Frederik T. Schut, "Assessing Hospital Competition When Prices Don't Matter to Patients: The Use of Time-Elasticities," *International Journal of Health Care Finance and Economics 10* (2010): 43–60.
Ventola, C. Lee, "Direct-to-Consumer Advertising: Therapeutic or Toxic," *P&T: Journal for Formulary Management 36* (2011): 669–684.
Vernon, John A., "Examining the Link between Price Regulation and Pharmaceutical R&D Investment," *Health Economics 14* (2005): 1–16.
Viscusi, W. Kip, "Age Variations in Risk Perceptions and Smoking Decisions," *Review of Economics and Statistics 73* (1991): 577–588.
Viscusi, W. Kip, "Cigarette Smokers and the Social Consequence of Smoking," *Tax Policy and the Economy 9* (1995): 51–101.
Viscusi, W. Kip, "Promoting Smokers' Welfare with Responsible Taxation," *National Tax Journal 57* (1994): 547–558.
Viscusi, W. Kip, "Risk Equity," *Journal of Legal Studies 29* (2000): 843–871.
Viscusi, W. Kip, "The Value of Risks to Life and Health," *Journal of Economic Literature 31* (1993): 1912–1946.
Viscusi, W. Kip, John M. Vernon, and Joseph E. Harrington, Jr., *Economics of Regulation and Antitrust*, 2nd ed. (Cambridge, MA: MIT Press, 1995).
Vistnes, Jessica P., and Brandy J. Lipton, "The Uninsured in America: Estimates of the Percentage of Non-Elderly Adults Uninsured throughout Each Calendar Year, by Selected Population Subgroups and State Medicaid Expansion Status: 2013 and 2014," Medical Expenditure Panel Survey, Statistical Brief #488, Washington, DC: Agency for Healthcare Research and Quality, June 2016.
Vitaliano, Donald F., "On the Estimation of Hospital Cost Functions," *Journal of Health Economics 6* (1987): 305–318.
Vitaliano, Donald F., and Mark Toren, "Cost Efficiency in Nursing Homes: A Stochastic Frontier Approach," *Journal of Health Economics 13* (1994): 281–300.
Vitaliano, Donald F., and Mark Toren, "Hospital Cost and Efficiency in a Regime of Stringent Regulation," *Eastern Economic Journal 22* (1996): 161–175.
Vitikainen, K., Linna, M., and Street, A., "Substituting Inpatient for Outpatient Care: What Is the Impact on Hospital Cost and Efficiency?" *European Journal of Health Economics 11* (2010): 395–404.
Waber, Rebecca L., et al., "Commercial Features of Placebo and Therapeutic Efficacy," *Journal of the American Medical Association 299* (2008): 1016–1017.
Wagstaff, Adam, "Estimating Efficiency in the Hospital Sector: A Comparison of Three Statistical Cost Frontier Models," *Applied Economics 21* (1989): 659–672.
Wagstaff, Adam, "QALYs and the Equity-Efficiency Trade-Off," *Journal of Health Economics 10* (1991): 21–41.

Wagstaff, Adam, and Guillem Lopez, "Hospital Costs in Catalonia: A Stochastic Frontier Analysis," *Applied Economics 3* (1996): 471–474.

Waldmann, Robert J., "Income Distribution and Infant Mortality," *Quarterly Journal of Economics 107* (1992): 1283–1302.

Wang, Jian, "Major Tasks for Five Priorities of Health Care Reform in 2011," China Center for Economic Research, manuscript, 2011.

Wang, Justin, et al., "Do Bad Report Cards Have Consequences?: Impacts of Publicly Reported Provider Quality Information on the CABG Market in Pennsylvania," *Journal of Health Economics 30* (2011): 392–407.

Warner, Kenneth E., and Linda M. Goldenhar, "The Cigarette Advertising Broadcast Ban and Magazine Coverage of Smoking and Health," *Journal of Public Health Policy 10* (1989): 32–42.

Warner, Kenneth E., et al., "Promotion of Tobacco Products: Issues and Policy Options," *Journal of Health Politics, Policy and Law 11* (1986): 367–392.

Webbink, D., N. Martin, and P. Visscher, "Does Education Reduce the Probability of Being Overweight?" *Journal of Health Economics 29* (2010): 29–38.

Wedig, Gerard J., "Health Status and the Demand for Health: Results on Price Elasticities," *Journal of Health Economics 7* (1988): 151–163.

Weiner, Jonathon P., "Forecasting the Effects of Health Reform on U.S. Physician Workforce Requirement," *Journal of the American Medical Association 272* (1994): 222–230.

Weisbrod, Burton A., "Costs and Benefits of Medical Research: A Case Study of Poliomyelitis," *Journal of Political Economy 79* (1971): 527–544.

Weisbrod, Burton A., *The Nonprofit Economy* (Cambridge, MA: Harvard, 1988).

Weisbrod, Burton A., "Toward a Theory of the Voluntary Non-Profit Sector in a Three-Sector Economy," in *Altruism, Morality, and Economic Theory*, Edmund S. Phelps (ed.) (New York: Russell Sage Foundation, 1975).

Wells, William, John Burnett, and Sandra Moriarity, *Advertising: Principles and Practice* (Upper Saddle River, NJ: Prentice Hall, 1989).

Wennberg, John E., "Dealing with Medical Practice Variations: A Proposal for Action," *Health Affairs 3* (1984): 6–32.

Wennberg, John E., "Population Illness Rates Do Not Explain Population Hospitalization Rates," *Medical Care 25* (1987): 354–359.

Wennberg, John E., "Small Area Analysis and the Medical Care Outcome Problem," in *Research Methodology: Strengthening Causal Interpretation of Non-Experimental Data*, E. Perrin and J. Bunker (eds.) Rockville: Department of Health and Human Services, 1990: 177–213.

Wennberg, John E., and F. J. Fowler, "A Test of Consumer Contribution to Small Area Variations in Health Care Delivery," *Journal of the Maine Medical Association 68* (1977): 275–279.

Wennberg, John E., and Alan Gittelsohn, "Small Area Variations in Health Care Delivery," *Science 182* (1973): 1102–1108.

Wennberg, John E., and Alan Gittelsohn, "Variations in Medical Care among Small Areas," *Scientific American 246* (1982): 120–134.

Wennberg, John E., B. A. Barnes, and M. Zubkoff, "Professional Uncertainty and the Problem of Supplier-Induced-Demand," *Social Science and Medicine 16* (1982): 811–824.

Wennberg, John E., et al., "Changes in Tonsillectomy Rates Associated with Feedback and Review," *Pediatrics 59* (1977): 821–826.

Wennberg, John E., et al., "Hospital Use and Mortality among Medicare Beneficiaries in Boston and New Haven," *New England Journal of Medicine 321* (1989): 1168–1173.

Wheeler, Benedict W., "Health-Related Environmental Indices and Environmental Equity in England and Wales," *Environment and Planning A 36* (2004): 803–822.

White, Chapin, James D. Reschovsky, and Amelia M. Bond, "Understanding Differences between High- and Low-Price Hospitals: Implication for Efforts to Rein in Costs," *Health Affairs 33* (2014): 324–331.

Whynes, David K., Emma J. Frew, and Jane L. Wolstenholme, "Willingness-to-Pay and Demand Curves: A Comparison of Results Obtained Using Different Elicitation Formats," *International Journal of Health Care Finance and Economics 5* (2005): 369–386.

Wickizer, Thomas M., and Paul J. Feldstein, "The Impact of HMO Competition on Private Health Insurance Premiums, 1985–92," *Inquiry 32* (1995): 241–251.

Wickizer, Thomas M., John R. C. Wheeler, and Paul J. Feldstein, "Have Hospital Inpatient Cost Containment Programs Contributed to the Growth in Outpatient Expenditures?" *Medical Care 29* (1991): 442–451.

Wiggins, Steven N., and Robert Mannes, "Price Competition in Pharmaceuticals: The Case of Anti-Infectives," *Economic Inquiry 42* (2004): 247–262.

Wilkinson, Richard, and Kate Pickett, *The Spirit Level: Why Greater Equality Makes Society Stronger* (London: Bloomsbury Press, 2009).

Williams, Alan, "The 'Fair Innings Argument' Deserves a Fairer Hearing!" *Health Economics 10* (2001): 583–585.

Williams, Alan, "Intergeneration Equity: An Exploration of the 'Fair Innings' Argument," *Health Economics 6* (1997): 117–132.

Williams, Alan, and Richard Cookson, "Equity in Health," *Handbook of Health Economics*, Volume I, A. J. Culyer and J. P. Newhouse (eds.) (Amsterdam: Elsevier Science, B.V., 2000).

Williams, Jenny, Lisa M. Powell, and Henry Wechsler, "Does Alcohol Consumption Reduce Human Capital Accumulation? Evidence from a College Alcohol Study," *Applied Economics 10* (2003): 1227–1239.

Williams, Scott C., et al., "Quality of Care in U.S. Hospitals as Reflected by Standardized Measures, 2002–2004," *New England Journal of Medicine 353* (2005): 255–264.

Wittman, Donald, "Efficient Rules in Highway Safety and Sports Activity," *American Economic Review 72* (1982): 78–90.

Wolfe, Barbara L., and Jere R. Behrman, "Women's Schooling and Children's Health: Are the Effects Robust with Adult Sibling Control for the Women's Childhood Background?" *Journal of Health Economics 6* (1987): 239–254.

Wolfe, John R., and John H. Goddeeris, "Adverse Selection, Moral Hazard, and the Wealth Effects in the Medigap Insurance Market," *Journal of Health Economics 10* (1991): 433–459.

Woolhandler, Steffie, Terry Campbell, and David U. Himmelstein, "Costs of Health Care Administration in the United States and Canada," *New England Journal of Medicine 349* (2003): 768–775.

World Bank Datasite, www.worldbank.org/data/databytopic/class.htm, accessed April 29, 2002.

World Health Organization, *Vital Statistics*, various years.

Writing Group for the Women's Health Initiative Investigators, "Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women," *Journal of the American Medical Association 288* (2002): 321–333.

Wu, Vivian Y., "Hospital Cost Shifting Revisited: New Evidence from the Balanced Budget Act of 1997," *International Journal of Health Care Finance and Economics 10* (2010): 61–83.

Wu, Vivian Y., "Managed Care's Price Bargaining with Hospitals," *Journal of Health Economics 28* (2009): 350–360.

Wynne, Billy, "May the Era of Medicare's Doc Fix (1997–2015) Rest in Peace. Now What?" *Health Affairs* blog, April 14, 2015, http://healthaffairs.org/blog/2015/04/14/may-the-era-of-medicares-doc-fix-1997–2015-rest-in-peace-now-what/, accessed February 25, 2016.

Yaisawarng, Suthathip, "Performance-based Budgeting in the Public Sector: An Illustration from the VA Health Care System," *Health Economics 15* (2006): 295–310.

Yaniv, Gideon, Odelia Rosin, and Yossef Tobel, "Junk-food, Home Cooking, Physical Activity and Obesity: The Effect of the Fat Tax and the Thin Subsidy," *Journal of Public Economics 93* (2009): 823–830.

Yilma, Zelalem, Luuk van Kempen, and Thomas de Hoop, "A Perverse 'Net' Effect? Health Insurance and Ex-Ante Moral Hazard in Ghana," *Social Science and Medicine 75* (2012): 138–147.

Yip, Winnie C., "Physician Response to Medicare Fee Reductions: Changes in the Volume of Coronary Bypass Graft (CABG) Surgeries in the Medicare and Private Sectors," *Journal of Health Economics 17* (1998): 675–699.

Yoo, Byung-Kwang, et al., "Impacts of Informal Caregiver Availability on Long-Term Care Expenditures in OECD Countries," *Health Services Research 39* (2004): 1971–1992.

Zaleski, Peter A., and Alfredo G. Esposito, "The Response to Market Power: Nonprofit Hospitals versus For-Profit Hospitals," *Atlantic Economic Journal 35* (2007): 315–325.

Zeckhauser, Richard J., and David V.P. Marks, "Sign-Posting Selected Product Information and Market Function," manuscript, John F. Kennedy School of Government, 1989.

Zhao, Lan, "Why Are Fewer Hospitals in the Delivery Business?" Working Paper 2007–04, The Walsh Center for Rural Health Analysis NORC, Health Policy & Evaluation Division, 2007.

Zivin, M., and M. Neidall, "Medical Technology Adoption, Uncertainty, and Irreversibilities: Is a Bird in the Hand Really Worth More than in the Bush?" *Health Economics 19* (2010): 142–153.

Zuckerman, Stephen, Jack Hadley, and Lisa Iezzoni, "Measuring Hospital Efficiency with Frontier Cost Functions," *Journal of Health Economics 13* (1994): 255–280.

Zuvekas, Samuel H., and Joel W. Cohen, "Fee-for-Service, While Much Maligned, Remains the Dominant Payment Method for Physician Visits," *Health Affairs 35* (2013): 411–414.

Zwanziger, Jack, and Cathleen Mooney, "Has Competition Lowered Prices?" *Inquiry 42* (2005): 73–85.

Author Index

Subject Index

Locators in *italics* refer to figures and those in **bold** to tables. The suffix g is used to indicate glossary entries.