Health Economics
Theory, Insights, and Industry Studies

Rexford E. Santerre
Stephen P. Neun

5th Edition
REXFORD E. SANTERRE
Professor of Finance and Healthcare Management
Programs in Healthcare Management and Insurance Studies
Department of Finance
School of Business
University of Connecticut

STEPHEN P. NEUN
Professor of Economics
Dean of Academic Affairs
Anna Maria College

SOUTH-WESTERN
CENGAGE Learning®

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States
Health Economics, 5e
Rexford E. Santerre and Stephen P. Neun
Vice President of Editorial, Business: Jack W. Calhoun
Publisher: Joe Sabatino
Acquisitions Editor: Steven Scoble
Supervising Developmental Editor: Katie Janos
Associate Marketing Manager: Betty Jung
Content Project Management: Pre-Press PMG
Media Editor: Deepak Kumar
Manufacturing Buyer: Sandee Milewski
Production Service: Pre-Press PMG
Copyeditor: Lorraine Martindale
Compositor: Pre-Press PMG
Senior Art Director: Michelle Kunkler
Cover Designer: Rose Alcorn
Cover Images: © Bogden Pop/iStockphoto; © Matthias Haas/iStockphoto; © Henrik Jonsson/iStockphoto

© 2010, 2007 South-Western, Cengage Learning
ALL RIGHTS RESERVED. No part of this work covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, tapeing, Web distribution, information storage and retrieval systems, or in any other manner—except as may be permitted by the license terms herein.

For product information and technology assistance, contact us at Cengage Learning Customer & Sales Support, 1-800-354-9706
For permission to use material from this text or product, submit all requests online at www.cengage.com/permissions
Further permissions questions can be emailed to permissionrequest@cengage.com

Library of Congress Control Number: 2009923659
ISBN-10: 0-324-78908-4

South-Western Cengage Learning
5191 Natorp Boulevard
Mason, OH 45040
USA

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For your course and learning solutions, visit www.cengage.com

Purchase any of our products at your local college store or at our preferred online store www.ichapters.com

Printed in the United States of America
1 2 3 4 5 6 7 8 12 11 10 09
To All the Girls in Our Lives:
Alexis
Cecilia
Jessica
Joan
Laurie
Lorraine
Mabel
Pamela
Patricia
Paula
Stephanie
Veronica
The health care sector, now representing more than one-sixth of the U.S. economy in terms of economic activity, continues to change in unimaginable ways. Sweeping transformations in the organizational arrangements of health care providers, newly developed medical technologies, the creation of new health insurance products, and the development and evaluation of various public policy initiatives all make the health care sector a dynamic and exciting area for applying the lens and tools of economic analysis. Indeed, not a day goes by without the unfolding of a medical event that requires the insights of economics to unravel the depths of its implications.

Our textbook, now in its fifth edition, is written expressly to capture the excitement generated by the health care field. As in the earlier editions, we take a fresh, contemporary approach to the study of health economics. We present the material in a lively and inviting manner by providing numerous and timely real-world examples throughout the text. At the same time, we resist the temptation of becoming overly encyclopedic and avoid purely technical issues that interest only academics and not students.

As a result of the approach taken, our book has wide appeal. Many business schools; liberal arts colleges; medical schools; and schools of public health, pharmacy, and health administration, at both the undergraduate and graduate levels, have chosen to use our textbook. In addition, the national Certified Employee Benefits Specialist (CEBS) program, cosponsored by the International Foundation of Employee Benefit Plans and The Wharton School of the University of Pennsylvania, has selected our text. The mix of adopters attests to the relevance and practicality of the material and the consistent and inviting manner in which various principles and concepts of health economics are presented throughout the text.

What’s New in the Fifth Edition?

In addition to updating all figures and adding new empirical studies, several changes have been made in this edition in response to the suggestions of various individuals.

- Among the more important changes, Chapter 1 now serves as more of an introductory chapter. To accomplish that objective, some material has been taken out and moved to other areas in the text. In the place of this previous material, a discussion has been added about the three legs of the medical care stool. The three legs of costs, access, and quality often act as barometers of a health economy. With that in mind we initially access the status of the U.S. health economy by presenting and examining time series data on national health care spending, along with its sources and uses of funds, uninsured rates, and infant mortality. We believe this new focus on the three-legged medical care stool acts as an effective springboard to motivate the rest of the material in the text.
Preface

- Appendix 1 provides material formerly in Chapter 1 on models and empirical estimation and provides a discussion on the difference between establishing association and causation when applying multiple regression analysis. This chapter also includes an introductory discussion about observational, quasi-experimental, instrumental, fixed effects, and social experiment studies.
- Chapter 2 offers more empirical evidence on factors influencing the health of infants and elderly individuals. The role of public health is also discussed in this chapter.
- Consumer-directed health plans are examined in Chapter 6.
- Chapter 8 now includes a discussion about measuring market power and the role of buyer power.
- The efficiency implication of a price ceiling is examined in greater detail in Chapter 9 and shown to depend on the type of moral hazard and competitiveness of the marketplace.
- Chapter 11 provides more information on \emph{ex post} and \emph{ex ante} moral hazard.
- Health insurance reform in Massachusetts is discussed in Chapter 16.

Organization of the Textbook

The textbook contains four parts: Part I, Chapters 1 through 8, deals with basic health economic concepts, such as trade-offs, the production of health, health care systems and institutions, the demands for medical care and health insurance, the health insurance product, production and cost theories, cost and benefit analysis, and market analysis. More specifically, Chapter 2 examines theoretically and empirically the different factors that help produce health. Not surprisingly, the role of medical care in producing health is given particular attention in this chapter.

Chapter 3 covers cost-benefit and cost effectiveness analysis, among other topics. Knowledge of these two methods helps policy makers determine efficient and effective ways to keep people healthy at minimum cost. An overview of health care system elements and an introduction to the U.S. health care system are provided in Chapter 4. A general model of a health care system and the role of financing, reimbursement, and delivery in a health economy are some of the issues discussed in this chapter.

Chapters 5 and 6 provide theoretical and empirical material on the demands for medical care and medical insurance. This information becomes important, for example, when asking questions concerning the utilization of medical care and why some people lack health insurance. Chapter 7 provides basic instruction on production and cost theories. These theories are crucial for understanding the behavior of any type of medical firm, regardless of its ownership type and how much competition it faces in the marketplace. Lastly, tools of market analysis are provided in Chapter 8. In this chapter, different market structures, such as perfect competition and monopoly, are discussed and compared in the context of a medical care industry.

In Part II, Chapters 9 and 10 focus on the important role of government in health matters and medical care markets. In particular, Chapter 9 provides an overview of government functions, such as regulation, antitrust, and redistribution, as applied to health and medical care issues. Chapter 10 discusses government’s ever-increasing role as a producer of health insurance and examines the Medicaid and Medicare programs in considerable detail.
Part III includes Chapters 11 through 15. These chapters use the concepts and theories developed in the earlier chapters to extensively analyze specific health care industries by applying the structure, conduct, and performance paradigm of industrial organization. The private health insurance, physician, hospital, pharmaceutical, and nursing home industries are covered in great depth, and the analysis is kept as current as possible. Various health care topics and issues are examined in these chapters.

Finally, Part IV, or Chapter 16, deals with health insurance reform. Some of the more debated plans for reforming the U.S. health insurance system at the federal and state levels are discussed and evaluated. The book ends with a glossary.

In most colleges and universities, a course in health economics is offered on a one-semester basis. Within one semester, it is difficult to cover all of the material in this text. The business curriculum at the University of Connecticut offers the typical health economics course in two semesters at both the undergraduate and MBA/MPH levels. (Not all students always take both courses, however.) The first-semester course is titled Health Insurance. This first course covers Chapters 4 (Health Care Systems and Institutions), 6 (The Demand for Medical Insurance), 10 (Government as Health Insurer), 11 (The Private Health Insurance Industry), and 16 (Health Care Reform). Parts of Chapter 2 (Health and Medical Care) are also covered before Chapter 6 and Chapter 8 (Structure, Conduct, Performance, and Market Analysis) is briefly reviewed before introducing Chapter 11.

The second-semester course is titled Health Care Economics, which covers Chapters 1 (Introduction), 2 (Health and Medical Care), 3 (Cost and Benefit Analysis), 5 (The Demand for Medical Services), 7 (Medical Care Production and Costs), 8 (Structure, Conduct, Performance, and Market Analysis), 9 (Government, Health, and Medical Care), and the four remaining industry chapters (12–15). Supplemental readings are assigned in both courses, and typically student presentations or point/counterpoint debates are assigned. Spreading the material over two courses means less rushing from topic to topic and provides more time to explore individual issues in greater detail. The students seem to appreciate the two-course approach.

Supplements

Economic Applications: Economic Applications includes South-Western’s dynamic Web features: EconNews, EconDebate, and EconData Online. Organized by pertinent economic topics and searchable by topic or feature, these features are easy to integrate into the classroom. EconNews, EconDebate, and EconData all deepen students’ understanding of theoretical concepts through hands-on exploration and analysis of the latest economic news stories, policy debates, and data. These features are updated on a regular basis. For more information, visit http://www.cengage.com/economics/infoapps

InfoTrac: With InfoTrac College Edition, students can receive anytime, anywhere online access to a database of full-text articles from thousands of popular and scholarly periodicals, such as Newsweek, Fortune, and Nation’s Business. InfoTrac is a great way to expose students to online research techniques, with the security that the content is academically based and reliable. For more information, visit http://www.cengage.com/economics/infoapps
Web Site: The support site for *Health Economics* can be accessed at www.cengage.com/economics/santerre and contains chapter-by-chapter web links, term paper tips, instructor resources, and other teaching and learning resources.

If a 1pass access card came with this book, you can start using many of these resources right away by following the directions on the card. One username and password gives you multiple resources. Get started today at www.cengage.com/economics/santerre/

Web-Based Instructor’s Manual: The *Health Economics* support site (www.cengage.com/economics/santerre) contains password-protected material for instructors only, including answers to end-of-chapter questions in the text, teaching notes for the case studies, a sample syllabus with web links, a list of readings for each chapter, and ideas for course projects.

PowerPoint™ Slides: PowerPoint slides are also located on the support site and are available for use by instructors for enhancing lectures. Each chapter’s slides include a lecture outline illustrated with key tables and graphs.

Instructor’s Resource CD-ROM: Get quick access to the Instructor’s Manual and PowerPoint slides from your desktop via one CD-ROM.

Acknowledgments

Our goal is to create the best possible learning device for students and teaching tool for professors. We are profoundly grateful to all of the reviewers for helping us bring this goal to fruition. For reviewing this fifth edition and providing numerous comments and suggestions, we thank Mir Ali, University of Toledo; Mark Barabas, Breyer State University; Jay Bhattacharya, Stanford University; David Bishai, Johns Hopkins University; Guy David, University of Pennsylvania; Derek DeLia, Rutgers; Ashley Hodgson, University of California-Berkeley; Timothy Leslie, George Mason University; Mindy Marks, University of California Riverside; Amalia Miller, University of Virginia; Gabriel Picone, University of South Florida.

We also appreciate the reviewers of past editions and others who have provided us with comments for improving the text over the years, including:

- Steven Andes
- Jay Bae
- Mary Ann Bailey
- Laurie Bates
- Richard Beil
- Sylvester Berki
- Bruce Carpenter
- Sewin Chan
- Chris Coombs
- Partha Deb
- Diane Dewar

- Randall Ellis
- Alfredo Esposto
- Maya Federman
- Andrew Foster
- A. Mark Freeman
- Linda Ghent
- Stephan Gohmann
- Glenn Graham
- Darren Grant
- Dennis Heffley
- Jim Hilliard

- Vivian Ho
- Robert Jantzen
- Juan Kelly
- Donald Kenkel
- Dong Li
- Frank Musgrave
- John Nyman
- Albert Oriol
- Irene Powell
- Keith Rayburn
- Jessica Wolpaw Reyes
As mentioned in the previous editions, if you have any comments or suggestions for improving the text, please bring them to our attention. We are only an email message away. We thank you in advance.

Rex Santerre
rsanterre@business.uconn.edu
Stephen Neun
sneun@annamaria.edu
## PART 1  BASIC HEALTH CARE ECONOMIC TOOLS AND INSTITUTIONS

### CHAPTER 1  Introduction

<table>
<thead>
<tr>
<th>What Is Health Economics?</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Four Basic Questions</td>
<td>5</td>
</tr>
<tr>
<td>Production and Allocative Efficiency and the Production Possibilities Curve</td>
<td>5</td>
</tr>
<tr>
<td>The Distribution Question</td>
<td>8</td>
</tr>
<tr>
<td>Implications of the Four Basic Questions</td>
<td>9</td>
</tr>
<tr>
<td>Taking the Pulse of the Health Economy</td>
<td>10</td>
</tr>
<tr>
<td>Medical Care Costs</td>
<td>10</td>
</tr>
<tr>
<td>Uses of Medical Funds</td>
<td>10</td>
</tr>
<tr>
<td>Sources of Medical Funds</td>
<td>11</td>
</tr>
<tr>
<td>Amount of Medical Care Spending</td>
<td>13</td>
</tr>
<tr>
<td>Medical Care Access</td>
<td>15</td>
</tr>
<tr>
<td>Medical Care Quality</td>
<td>16</td>
</tr>
<tr>
<td>A Note on the Relation between System Structure and Performance</td>
<td>17</td>
</tr>
<tr>
<td>Summary</td>
<td>19</td>
</tr>
<tr>
<td>Review Questions and Problems</td>
<td>19</td>
</tr>
<tr>
<td>References</td>
<td>21</td>
</tr>
<tr>
<td><strong>Appendix 1: Economic Models and Empirical Testing</strong></td>
<td>21</td>
</tr>
<tr>
<td>Economic Models</td>
<td>21</td>
</tr>
<tr>
<td>Positive and Normative Analysis</td>
<td>24</td>
</tr>
<tr>
<td>Empirical Testing</td>
<td>25</td>
</tr>
<tr>
<td>Association versus Causation</td>
<td>29</td>
</tr>
<tr>
<td>Summary</td>
<td>32</td>
</tr>
<tr>
<td>Review Questions and Problems</td>
<td>33</td>
</tr>
<tr>
<td>References</td>
<td>35</td>
</tr>
</tbody>
</table>

### CHAPTER 2  Health and Medical Care: An Economic Perspective

| What Is Health?               | 38 |
| Why Good Health? Utility Analysis | 38 |
| What Is Medical Care?         | 41 |
| The Production of Good Health | 42 |
| Empirical Evidence on the Production of Health in the United States | 49 |
| The Determinants of Health among Nonelderly Adults | 50 |
| The Determinants of Health among Children | 53 |
| The Determinants of Health among the Elderly | 55 |
| The Role of Public Health: An Historical Approach | 56 |
| The Ten Major Causes of Death in the United States in 2005 | 57 |
| Empirical Evidence on the Production of Health: A Summary | 59 |
| Summary                      | 59 |
| Review Questions and Problems | 60 |
| Online Resources              | 61 |
| References                    | 62 |
### CHAPTER 3 Cost and Benefit Analysis

- Cost Identification Analysis 66
- Cost-Benefit Analysis 67
  - The Practical Side of Using Cost-Benefit Analysis to Make Health Care Decisions 71
  - Discounting 72
  - The Value of Life 73
  - An Application of Cost-Benefit Analysis—Should College Students Be Vaccinated? 76
  - The Costs and Benefits of New Medical Technologies 77
- Cost-Effectiveness Analysis 79
  - By the Numbers: Cost-Effectiveness and Cost-Utility Analysis 83
  - An Application of Cost-Effectiveness Analysis: Autologous Blood Donations—Are They Cost Effective? 84
- Summary 85
- Review Questions and Problems 86
- Online Resources 88
- References 88

### CHAPTER 4 Health Care Systems and Institutions

- Elements of a Health Care System 92
- The Role and Financing Methods of Third-Party Payers 94
- Risk Management, Reimbursement, and Consumer Cost Sharing 96
- The Production of Medical Services 99
  - Institutional Differences between For-Profit and Not-for-Profit Health Care Providers 100
  - Why Are Not-for-Profit Health Care Providers So Prevalent? 100
- Production of Health Care in the Three Systems 101
- Physician Choice and Referral Practices 102
- The Three National Health Care Systems Summarized 102
- An Overview of the U.S. Health Care System 104
- Financing of Health Care in the United States 104
- Reimbursement for Health Care in the United States 105
- Production of Health Services and Provider Choice in the United States 106
- Summary 107
- Review Questions and Problems 107
- Online Resources 108
- References 108

### CHAPTER 5 The Demand for Medical Care

- The Demand for Medical Care and the Law of Demand 110
  - The Utility-Maximizing Rule 110
  - The Law of Demand 112
  - Other Economic Demand-Side Factors 113
  - The Relationship between Health Insurance and the Demand for Medical Care 115
  - Moral Hazard 120
  - Noneconomic Determinants of the Demand for Medical Care 120
  - The Market Demand for Medical Care 122
  - The Fuzzy Demand Curve 123
  - Elasticities 124
  - Own-Price Elasticity of Demand 124
  - Other Types of Elasticity 128
  - Empirical Estimation 128
    - Own-Price, Income, Cross-Price, and Time-Cost Elasticity Estimates 129
    - The Impact of Insurance on the Demand for Medical Care 132
<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Impact of Noneconomic Factors on the Demand for Medical Services</td>
</tr>
<tr>
<td>Summary</td>
</tr>
<tr>
<td>Review Questions and Problems</td>
</tr>
<tr>
<td>Online Resources</td>
</tr>
<tr>
<td>References</td>
</tr>
</tbody>
</table>

**CHAPTER 6** The Demand for Medical Insurance: Traditional and Managed Care Coverage  
Introduction | 141 |
| The Conventional Theory of the Demand for Private Health Insurance | 142 |
| Deriving the Demand for Private Health Insurance | 143 |
| Factors Affecting the Quantity Demanded of Health Insurance | 144 |
| Nyman’s Access Theory of the Demand for Private Health Insurance | 145 |
| Conventional Insurance Theory According to Nyman | 146 |
| A Simple Exposition of the Nyman Model | 147 |
| Insights and Policy Implications of the Nyman Model | 148 |
| The Health Insurance Product: Traditional versus Managed Care Insurance | 149 |
| Financial Incentives and Management Strategies Facing Consumers/Patients | 150 |
| Financial Incentives and Management Strategies Facing Health Care Providers | 151 |
| Consumer-Directed Health Care Plans | 152 |
| The Regulation of MCOs | 153 |
| Summary | 155 |
| Review Questions and Problems | 156 |
| Online Resources | 157 |
| References | 157 |

**CHAPTER 7** Medical Care Production and Costs  
The Short-Run Production Function of the Representative Medical Firm | 177 |
| Marginal and Average Products | 178 |
| Elasticity of Input Substitution | 179 |
| A Production Function for Hospital Admissions | 180 |
| Short-Run Cost Theory of the Representative Medical Firm | 181 |
| The Short-Run Cost Curves of the Representative Medical Firm | 182 |
| Short-Run Per-Unit Costs of Production | 183 |
| Factors Affecting the Position of the Short-Run Cost Curves | 184 |
| Estimating a Short-Run Cost Function for Hospital Services | 185 |
| The Cost-Minimizing Input Choice | 186 |
| Long-Run Costs of Production | 187 |
| Long-Run Cost Curves | 188 |
| Shifts in the Long-Run Average Cost Curve | 189 |
| Long-Run Cost Minimization and the Indivisibility of Fixed Inputs | 190 |
| Neoclassical Cost Theory and the Production of Medical Services | 191 |
| Summary | 192 |
| Review Questions and Problems | 193 |
| Online Resources | 194 |
| References | 194 |

**CHAPTER 8** Structure, Conduct, Performance, and Market Analysis  
Structure, Conduct, and Performance Paradigm | 207 |
| Is a Perfectly Competitive Market Relevant to Medical Care? | 208 |
| A Model of Supply and Demand | 209 |
| Comparative Static Analysis | 210 |
Contents

A Note on Long-Run Entry and Exit in a Perfectly
Competitive Market 216
Using Supply and Demand to Explain Rising Health Care Costs 217
The Monopoly Model of Market Behavior and Performance 218
Monopoly versus Perfect Competition 219
Barriers to Entry 220
The Buyer Side of the Market 223
Monopolistic Competition and Product Differentiation 224
Procompetitive and Anticompetitive Aspects of Product
Differentiation 226
Oligopoly 228
Collusive Oligopoly 228
Competitive Oligopoly 229
Collusive or Competitive Oligopoly? 230
Oligopolistic Behavior in Medical Care Markets 231
Defining the Relevant Market, Measuring Concentration, and
Identifying Market Power 232
The Relevant Product and Geographical Markets 233
Measuring Market Concentration 234
Identifying Market Power 236
Summary 238
Review Questions and Problems 240
Online Resources 243
References 243

PART 2  THE ROLE OF GOVERNMENT 245

CHAPTER 9 Government, Health, and Medical Care 247
Economic Reasons for Government Intervention 248
Types of Government Intervention 250
Public Goods 251
Externalities 251
Regulations 261
Antitrust Laws 268
Public Enterprise 275
The Redistribution Function of Government 276
Summary 280
Review Questions and Problems 281
Online Resources 282
References 282

CHAPTER 10 Government as Health Insurer 285
Why Does the Government Produce Health Insurance? 286
The Medicaid Program 287
The Financing and Cost of Medicaid 288
Do Differences in the Medicaid Program Make a Difference? 289
The State Children’s Health Insurance Program 291
Are the Medicaid and State Children’s Health Insurance
Programs “Crowding Out” Private Health Insurance? 292
The Medicare Program 293
The Financing and Cost of Medicare 294
Medicare Program Reforms 296
Medicare Prescription Drug Improvement and
Modernization Act of 2003 306
Summary 309
Review Questions and Problems 310
PART 3   INDUSTRY STUDIES   313

CHAPTER 11   The Private Health Insurance Industry   315
A Brief History of the Private Health Insurance Industry  316
The Structure of the Private Health Insurance Industry  317
  Number, Types, and Size Distribution of Health Insurers  317
  Barriers to Entry  320
  Consumer Information  321
The Conduct of the Private Health Insurance Industry  322
  The Dominant Insurer Pricing Model  323
  Other Issues Relating to the Pricing of Health Insurance  326
  Managed Care Organizations and Insurance Premiums  327
  Do HMOs Possess Monopsony Power?  329
  Rating of Premiums, Adverse Selection, and Risk Selection  332
  Guaranteed Renewability in the Individual Health Insurance Market  336
The Performance of the Private Health Insurance Industry  337
  The Price of Private Health Insurance in the United States  338
  Output of Private Health Insurance in the United States  344
  Profitability in the Private Health Insurance Industry  350
Summary  352
Review Questions and Problems  352
Online Resources  354
References  354

CHAPTER 12   The Physician Services Industry   357
The Structure of the Physician Services Industry  358
  The Number of Physicians in the United States  358
  Distribution of Primary Care and Specialty Care Physicians in the United States  360
  Mode of Practice  362
  Buyers of Physician Services and Methods of Remuneration  363
  Reimbursement Practices of Managed Care Buyers of Physician Services  364
  Barriers to Entry  364
  Production, Costs, and Economies of Scale  367
Summary of the Structure of the Market for Physician Services  369
The Conduct of the Physician Services Industry  369
  The Supplier-Induced Demand Hypothesis  370
  McGuire’s Quantity-Setting Model  372
  The Impact of Alternative Compensation Schemes on Physician Behavior  375
  Geographical Variations in the Utilization of Physician Services  377
  The Impact of Utilization Review on the Physician Services Market  379
  Medical Negligence and Malpractice Insurance  381
The Performance of the Physician Services Industry  386
  Expenditures on Physician Services  386
  The Physician Services Price Inflation Rate  386
  The Utilization of Physician Services  388
  Physician Income  388
Summary  390
Review Questions and Problems  390
Online Resources  391
References  391
### CHAPTER 13 The Hospital Services Industry

- **The Structure of the Hospital Services Industry** 398
  - Number, Types, and Size Distribution of U.S. Hospitals 398
  - Measuring Market Concentration 399
  - Barriers to Entry 401
  - Number, Types, and Size Distribution of the Buyers of Hospital Services 406
  - Type of Product 408
- **Summary of the Structure of the Hospital Services Industry** 411

- **The Conduct of the Hospital Services Industry** 412
  - Pricing Behavior of Not-for-Profit Hospitals 412
  - Inside the Black Box of the Hospital 419
  - Market Concentration and Hospital Behavior 420
  - Hospital Ownership and Hospital Behavior 423
  - Managed-Care Buyers and Hospital Behavior 425
  - Price Regulations and Hospital Behavior 428
  - Cost Shifting Behavior 430
  - Integrated Delivery Systems 432
- **Summary of the Conduct of the Hospital Services Market** 437

- **The Performance of the Hospital Services Industry** 438
  - The Growth in Hospital Expenditures 438
  - The Hospital Services Price Inflation Rate 439
  - Hospital Input Usage and Utilization 441
  - Hospital Profit Margins 444
- **Summary** 446

### CHAPTER 14 The Pharmaceutical Industry

- **The Structure of the Pharmaceutical Industry** 458
  - Number and Size Distribution of Sellers 458
  - The Buyer Side of the Pharmaceutical Market 461
  - Barriers to Entry 466
  - Consumer Information and the Role of the FDA 468
- **The Structure of the Pharmaceutical Industry: A Summary** 470

- **The Conduct of the Pharmaceutical Industry** 470
  - Pricing Behavior 471
  - Promotion of Pharmaceutical Products 474
  - Product Innovation 476
- **The Conduct of the Pharmaceutical Industry: A Summary** 482

- **The Performance of the Pharmaceutical Industry** 482
  - The Relative Price Inflation Rate of Pharmaceutical Products 483
  - Output of New Pharmaceutical Products 484
  - Profits in the Pharmaceutical Industry 488
- **The Performance of the Pharmaceutical Industry: A Summary** 490

### CHAPTER 15 The Long-Term Care Industry

- **The Structure of the Long-Term Care Services Industry** 498
  - The Need for Long-Term Care 499
  - Structure of Informal Care Providers 500
<table>
<thead>
<tr>
<th>Contents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure of the Nursing Home Care Industry</td>
<td>501</td>
</tr>
<tr>
<td>Structure of the Home Health Care and Hospice Industry</td>
<td>505</td>
</tr>
<tr>
<td>The Structure of the Long-Term Care Industry: A Summary</td>
<td>507</td>
</tr>
<tr>
<td>The Conduct of the Long-Term Care Industry</td>
<td>507</td>
</tr>
<tr>
<td>The Dual Market Model of Nursing Home Pricing</td>
<td>507</td>
</tr>
<tr>
<td>The Effect of Alternative Payment Methods</td>
<td>511</td>
</tr>
<tr>
<td>Scale Economies with Respect to Quality</td>
<td>512</td>
</tr>
<tr>
<td>Ownership and Conduct</td>
<td>513</td>
</tr>
<tr>
<td>Market Concentration and Nursing Home Conduct</td>
<td>516</td>
</tr>
<tr>
<td>Conduct of the Long-Term Care Industry: A Summary</td>
<td>519</td>
</tr>
<tr>
<td>The Performance of the Long-Term Care Industry</td>
<td>521</td>
</tr>
<tr>
<td>Expenditures on Long-Term Care</td>
<td>521</td>
</tr>
<tr>
<td>Formal Expenditures on Long-Term Care</td>
<td>522</td>
</tr>
<tr>
<td>Private Insurance for Long-Term Care</td>
<td>525</td>
</tr>
<tr>
<td>Prices for Nursing Home Services</td>
<td>527</td>
</tr>
<tr>
<td>The Utilization of Long-Term Care Facilities</td>
<td>528</td>
</tr>
<tr>
<td>What Do the Demographics Tell Us about the Future of Long-Term Care?</td>
<td>530</td>
</tr>
<tr>
<td>Summary</td>
<td>532</td>
</tr>
<tr>
<td>Review Questions and Problems</td>
<td>533</td>
</tr>
<tr>
<td>Online Resources</td>
<td>535</td>
</tr>
<tr>
<td>References</td>
<td>535</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PART 4 HEALTH INSURANCE REFORM</th>
<th>539</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 16 Health Insurance Reform</td>
<td>541</td>
</tr>
<tr>
<td>The Performance of the U.S. Health Care System:</td>
<td></td>
</tr>
<tr>
<td>A Summary and an International Comparison</td>
<td>542</td>
</tr>
<tr>
<td>Summarizing the Structure, Conduct, and Performance of the Various Medical Care Markets</td>
<td>542</td>
</tr>
<tr>
<td>The U.S. Health Care Economy: An International Comparison</td>
<td>547</td>
</tr>
<tr>
<td>Why Is There So Much Disagreement Concerning</td>
<td></td>
</tr>
<tr>
<td>How Health Care And Insurance Reform Should Be Designed?</td>
<td>551</td>
</tr>
<tr>
<td>An Overview of Health Insurance Reform in the United States</td>
<td>553</td>
</tr>
<tr>
<td>Health Insurance Reform at the Federal Level</td>
<td>554</td>
</tr>
<tr>
<td>Managed Competition</td>
<td>556</td>
</tr>
<tr>
<td>National Health Insurance</td>
<td>559</td>
</tr>
<tr>
<td>Medical Savings Accounts</td>
<td>560</td>
</tr>
<tr>
<td>Individual Mandates</td>
<td>561</td>
</tr>
<tr>
<td>Attempts at State Health Care and Insurance Reform:</td>
<td></td>
</tr>
<tr>
<td>Successes and Failures</td>
<td>562</td>
</tr>
<tr>
<td>Hawai'i: The Case of Employer Mandates</td>
<td>563</td>
</tr>
<tr>
<td>Maryland: The Case of Regulation</td>
<td>564</td>
</tr>
<tr>
<td>Massachusetts: The Case of Individual Mandates</td>
<td>564</td>
</tr>
<tr>
<td>Oregon: The Case of Rationing Medicaid Services</td>
<td>565</td>
</tr>
<tr>
<td>What Can Be Gained from State Attempts at Health Care Reform?</td>
<td>567</td>
</tr>
<tr>
<td>An Overview of the Clinton Health Care Plan</td>
<td>567</td>
</tr>
<tr>
<td>The Collapse of the Clinton Health Care Plan</td>
<td>568</td>
</tr>
<tr>
<td>Summary</td>
<td>569</td>
</tr>
<tr>
<td>Review Questions and Problems</td>
<td>570</td>
</tr>
<tr>
<td>Online Resources</td>
<td>571</td>
</tr>
<tr>
<td>References</td>
<td>571</td>
</tr>
</tbody>
</table>
PART ONE

Basic Health Care Economic Tools and Institutions

CHAPTER 1  Introduction
2 Health and Medical Care: An Economic Perspective
3 Cost and Benefit Analysis
4 Health Care Systems and Institutions
5 The Demand for Medical Care
6 The Demand for Medical Insurance: Traditional and Managed Care Coverage
7 Medical Care Production and Costs
8 Structure, Conduct, Performance, and Market Analysis
Introduction

Like millions of Americans at some point in their lives, Joe awoke one night feeling a crushing weight on his chest. As the pain spread down his arm, he realized he was experiencing his worst dread: a heart attack. His wife, Angela, called the paramedics. While the ambulance rushed Joe to the hospital, she anguished over the kind of care he would receive. Angela’s anxiety starkly illustrates the basic questions any health care system faces:

1. Who should receive medical goods and services? Would a person like Joe receive care merely because he is a citizen, or would he receive care only if he worked for a large company that provides health insurance for its employees?
2. What types of medical goods and services should be produced? Should the most expensive tests (such as angiograms) be performed without regard to cost? What treatments (such as balloon angioplasties) should be provided?
3. What inputs should be used to produce medical goods and services? Should the hospital use high-tech medical equipment, a large nursing staff, or both?¹

All health care systems face questions such as these, but sometimes choose to answer them differently. When responding to health and health care questions, societies around the world take into account important moral, cultural, legal, economic, and other considerations. Addressing all of these concerns simultaneously and thoroughly is a daunting task, in part, because one concern often conflicts with another, but also because this task involves a substantial amount of time, effort, and knowledge. Indeed, the intellectual resource commitment would be so great that no one book could adequately cover all of the pertinent issues.

Instead, this textbook focuses solely on the economic aspects of questions involving health and health care. The general objective of this textbook is to develop a set of analytical and conceptual tools that can be used to gain valuable insights into a host of health care issues and problems from an economic perspective. This chapter takes the first step in accomplishing this important objective. In particular, this chapter:

• introduces the discipline of health economics
• discusses resource constraints, trade-offs, efficiency, and equity
• highlights the state of the health economy in the United States and sets the stage for the material in the remaining chapters

¹We are indebted to Gary Wyckoff of Hamilton College for providing us with this example.
What Is Health Economics?

For many of you, this textbook provides your first exposure to the study of health economics. Perhaps the ongoing controversy regarding health care reform or the prospect of a career in the health care field motivated you to learn more about health economics. Or perhaps you need only three more credits to graduate. Whatever the reason, we are sure you will find health economics to be challenging, highly interesting, and personally rewarding.

The study of health economics involves the application of various microeconomics tools, such as demand or cost theory, to health issues and problems. The goal is to promote a better understanding of the economic aspects of health care problems so that corrective health policies can be designed and proposed. A thorough understanding of microeconomic analysis is essential for conducting sound health economics analyses. If you lack a background in microeconomics, don’t worry. This textbook is intended to help you learn and apply basic microeconomic theory to health economics issues. Before long, you will be thinking like a health economist!

The tools of health economics can be applied to a wide range of issues and problems pertaining to health and health care. For example, health economics analysis might be used to investigate why 25 of every 1,000 babies born in Turkey never reach their first birthday, whereas all but 3 of every 1,000 babies born in Japan live to enjoy their first birthday cake. The tools of health economics analysis might also be used to examine the economic desirability of a hotly contested merger between two large hospitals in a major metropolitan area. The burning question is: Will the merger of the two hospitals result in lower hospital prices due to overall cost savings or higher prices due to market power?

Health economics is difficult to define in a few words because it encompasses such a broad range of concepts, theories, and topics. The Mosby Medical Encyclopedia (1992, p. 361) defines health economics as follows:

Health economics . . . studies the supply and demand of health care resources and the impact of health care resources on a population.

Notice that health economics is defined in terms of the determination and allocation of health care resources. This is logical, because medical goods and services cannot exist without them. Health care resources consist of medical supplies, such as pharmaceutical goods, latex rubber gloves, and bed linens; personnel, such as physicians and lab assistants; and capital inputs, including nursing home and hospital facilities, diagnostic and therapeutic equipment, and other items that provide medical care services. Unfortunately, health care resources, like resources in general, are limited or scarce at a given point in time, and wants are limitless. Thus, trade-offs are inevitable and a society, whether it possesses a market-driven or a government-run health care system, must make a number of fundamental but crucial choices. These choices are normally couched in terms of four basic questions, discussed next.

---

2. Even health care services produced in the home, such as first aid (therapeutic services) or home pregnancy tests (diagnostic services), require resources.
The Four Basic Questions

As just noted, resources are scarce. Scarcity means that each society must make important decisions regarding the consumption, production, and distribution of goods and services as a way of providing answers to the four basic questions:

1. What mix of nonmedical and medical goods and services should be produced in the macroeconomy?
2. What mix of medical goods and services should be produced in the health economy?
3. What specific health care resources should be used to produce the chosen medical goods and services?
4. Who should receive the medical goods and services that are produced?

How a particular society chooses to answer these four questions has a profound impact on the operation and performance of its health economy.

The first two questions deal with allocative efficiency: What is the best way to allocate resources to different consumption uses? The first decision concerns what combination of goods and services to produce in the overall economy. Individuals in a society have unlimited wants regarding nonmedical and medical goods and services, yet resources are scarce. As a result, decisions must be made concerning the best mix of medical and nonmedical goods and services to provide, and this decision-making process involves making trade-offs. If more people are trained as doctors or nurses, fewer people are available to produce nonmedical goods such as food, clothing, and shelter. Thus, more medical goods and services imply fewer nonmedical goods and services, and vice versa, given a fixed amount of resources.

The second consumption decision involves the proper mix of medical goods and services to produce in the health economy. This decision also involves trade-offs. For example, if more health care resources, such as nurses and medical equipment, are allocated to the production of maternity care services, fewer resources are available for the production of nursing home care for elderly people. Allocative efficiency in the overall economy and the health economy is achieved when the best mix of goods is chosen given society’s underlying preferences.

The third question—what specific health care resources should be used?—deals with production efficiency. Usually resources or inputs can be combined to produce a particular good or service in many different ways. For example, hospital services can be produced in a capital- or labor-intensive manner. A large amount of sophisticated medical equipment relative to the number of patients served reflects a capital-intensive way of producing hospital services, whereas a high nurse-to-patient ratio indicates a labor-intensive process. Production efficiency implies that society is getting the maximum output from its limited resources because the best mix of inputs has been chosen to produce each good.

Production and Allocative Efficiency and the Production Possibilities Curve

The most straightforward way to illustrate production and allocative efficiency is to use the production possibilities curve (PPC). A PPC is an economic model that depicts the various combinations of any two goods or services that can be produced efficiently given the stock of resources, technology, and various institutional arrangements. Figure 1–1 displays
We assume society has already made its choice between medical and nonmedical goods. A PPC. The quantities of maternity services, $M$, and nursing home services, $N$, are shown on the vertical and horizontal axes, respectively. Points on the bowed-out PPC depict the various combinations of maternity and nursing home care services that can be efficiently produced within a health economy assuming the amounts of health care resources and technology are fixed at a given point in time.

Every point on the PPC implies production efficiency, since all health care resources are being fully utilized. For example, notice points $A$, $B$, $C$, $D$, and $E$ on the PPC. At each of these points, medical inputs are neither unemployed nor underemployed (for example, a point in the interior, such as $F$, reflects inefficiency because more of one good can be attained without necessarily reducing the other. A point outside the PPC, such as $G$, is not yet attainable but can be reached with an increase in resources or through institutional or technological changes that improve productivity.

3. We assume society has already made its choice between medical and nonmedical goods.
nurse involuntarily working part time rather than full time) and are being used in the most productive manner so that society is getting their maximum use. If a movement along the curve from one point to another occurs, units of one medical service must be forgone to receive more units of the other medical service.

Specifically, assume the health economy is initially operating at point C with \( M_C \) units of maternity care services and \( N_C \) units of nursing home services. Now suppose health care decision makers decide that society is better off at point D with one more unit of nursing home services, \( N_D - N_C \). The movement from point C to point D implies that \( M_C - M_D \) units of maternity care services are given up to receive the additional unit of nursing home services. Because medical resources are fully utilized at point C, a movement to point D means that medical inputs must be drawn or reallocated from the maternity care services market to the nursing home services market. As a result, the quantity of maternity care services must decline if an additional unit of nursing home services is produced. The forgone units of maternity care services, \( M_C - M_D \), represent the opportunity cost of producing an additional unit of nursing home services.4 Generally, opportunity cost is the value of the next best alternative that is given up.

The bowed-out shape of the PPC implies that opportunity cost is not constant but increases with a movement along the curve. Imperfect substitutability of resources is one reason for this so-called law of increasing opportunity cost. For example, suppose the nursing home services market expands downward along the PPC. To produce more nursing home services, employers must bid resources away from the maternity care services market. Initially, the least productive inputs in the maternity care services market are likely to be bid away, because they are available at a lower cost to nursing home employers. Consequently, very few maternity care services are given up at first. As the nursing home services market continues to expand, however, increasingly productive inputs in the maternity care services market must be drawn away. The implication is that society gives up ever-increasing units of maternity care services. Thus, the law of increasing opportunity cost suggests that ever-increasing amounts of one good must be given up to receive successively more equal increments of another good.

If medical inputs are not fully utilized because some inputs are idle or used unproductively, more units of one medical service can be produced without decreasing the amount of the other medical service. An example of an underutilization of resources is indicated by point F in the interior of the PPC. At point F, the health care system is producing only \( M_F \) units of maternity services and \( N_F \) units of nursing home services. Notice that by moving to point B on the PPC, both maternity care services and nursing home services can be increased without decreasing the other. The quantities of both goods increase only because some resources are initially idle or underutilized at point F. Health care resources are inefficiently employed at point F.

A point outside the current PPC, such as G, is attainable in the future if the stock of health care resources increases; a new, productivity-enhancing technology is discovered; or various economic, political, or legal arrangements change and improve productive relationships in the health economy. If so, the PPC shifts out and passes through a point like G. For example, technological change may enable an increased production of both maternity and

---

4. As economists are fond of reminding noneconomists, “There is no such thing as a free lunch!”
nursing home services from the same original stock of health care resources. Alternatively, a greater quantity of maternity and nursing home services can be produced and the PPC shifts outward if more people enter medical professions (possibly at the expense of all other goods and services).

Production efficiency is attained when the health economy operates at any point on the PPC, since medical inputs are producing the maximum amount of medical services and no unproductive behavior or involuntary unemployment exists. Allocative efficiency is attained when society chooses the best or most preferred point on the PPC. All points on the PPC are possible candidates for allocative efficiency. The ideal, or optimal, point for allocative efficiency depends on society’s underlying preferences for the two medical services.

Of course, the real world is much more complex than the example depicted by the PPC. Rather than only two goods, an unimaginable number of goods and services are produced in a society. The PPC is a model because it offers a simplification of reality. As pointed out in more depth in appendix A1, models are useful in the field of economics because they serve as conceptual devices or tools for organizing our thoughts about a topic. The PPC provides a good example of a simple but powerful model because it sheds light on a number of important lessons including: (1) the all-important economic role of scarcity; (2) the significance of economic choices; (3) the costs of inefficiency; and (4) how growth takes place in an economy.

The Distribution Question

The answer to the fourth question—who should receive the medical goods and services?—deals with distributive justice or equity. It asks whether the distribution of services is equitable, or fair, to everyone involved. In practice, countries around the world have chosen to address this medical care distribution question in many different ways.

When thinking about the distribution question, it is sometimes useful to consider two theoretically opposite ways of distributing output: the pure market system and a perfect egalitarian system. Goods and services are distributed in a pure market system based solely on each person’s willingness and ability to pay because decisions concerning the four basic questions are answered on a decentralized basis within a system of markets. That is, goods and services are distributed, or rationed, to only those people who are both willing and able to purchase them in the marketplace. Because people face an incentive to earn income to better afford goods and services in a pure market system, they tend to work hard and save appropriately for present and future consumption. Consequently, productive resources tend to be allocated efficiently in a pure market system. In other words, the incentives associated with a pure market system typically mean that the economy operates on the PPC.

In many cases, differences in ability to pay among individuals reflect that some have consciously chosen to work harder and save more than others. Unfortunately, differences in ability to pay may also indicate that some people have less income because of unfortunate life circumstances such as a mental, physical, or social limitation. Regardless of the specific reason, it follows that people without sufficient incomes face a financial barrier to obtaining goods and services in a pure market system in which price serves as a rationing mechanism. Given income disparities, some people may be denied access to needed goods and services. Consequently, the pure market system is typically viewed as inherently unfair by many when it comes to the distribution of important goods and services such as health care.
In direct contrast, a central committee, such as a federal or subnational unit of government, may answer the distribution question by ensuring everyone receives an equal share of goods and services. In an egalitarian system of this kind, everyone has access to the same goods and services without regard to income status or willingness to pay. Therefore, no one is denied access to needed goods and services. But an incentive may exist for people to choose to work and save less because the consumption decision is divorced from the distribution of earned income. Because of this inefficient allocation of resources, fewer goods and services may be available for distribution in an egalitarian system. In this case, the economy may operate inside the PPC.

In practice, most countries have adopted a mixed distribution system, with the reliance on central versus market distribution varying by degree across countries. For example, in the United States, many goods and services are distributed by both the market and the government. The food stamp, temporary assistance for needy families, and Medicaid programs represent some of the many policies adopted by the U.S. government to redistribute goods and services. Some people applaud these programs, whereas others argue that they worsen both efficiency and equity. They argue that efficiency and equity are compromised when those who choose to commit fewer resources to production are rewarded through redistributive programs and productive individuals are penalized via taxation. The efficiency and equity implications of various redistributive policies are constantly debated in the United States and elsewhere. In the context of health care, the consequence of this debate regarding distribution might determine who lives and who dies. For this reason, among others, more discussion on the redistributive function of government is taken up in Chapters 9 and 10.

Implications of the Four Basic Questions

Given a scarcity of economic resources, a society generally wishes to produce the best combination of goods and services by employing least-cost methods of production. Trade-offs are inevitable. As the PPC illustrates, some amount of one good or service must be given up if the production and consumption of another good or service increases. As a result, each society must make hard choices concerning consumption and production activities because scarcity exists. Choices may involve sensitive trade-offs, for example, between the young and the old, between prevention and treatment, or between men (prostate cancer) and women (breast cancer).

In addition, some individuals lack financial access to necessary goods and services such as food, housing, and medical care. Because achieving equity is a desirable goal, a society usually seeks some redistribution of income. Normally, the redistribution involves taxation. However, a tax on labor or capital income tends to create a disincentive for employing resources in their most efficient manner. Inefficient production suggests that fewer goods and services are available in the society (production inside the PPC). Thus, a trade-off often exists between equity and efficiency goals, and, consequently, hard choices must be made between the two objectives. The design of a nation’s health care system normally reflects the way the society has chosen to balance efficiency and equity concerns.

5. This point is discussed in more detail in Chapter 9.
Taking the Pulse of the Health Economy

A health economy, like a macroeconomy, involves the production and consumption of goods and services and the distribution of those goods to consumers. A health economy differs from a macroeconomy because it distinctly considers production, consumption, and distribution activities that directly relate to population health. More will be said about that difference in chapters 2 and 4. Another difference concerns the way in which economists take the pulse of the macroeconomy and health economy. While economists are really concerned with efficiency and equity, the unemployment, inflation, and gross domestic product growth rates are also considered when gauging the performance of a macroeconomy. If you recall from ECON 100, gross domestic product (GDP) captures the total market value of all goods and services produced in an economy during a particular period.

For a health economy, the analogous performance indicators are the components that make up the so-called three-legged stool of medical care: costs, access, and quality. Again, although health economists are more concerned about efficiency and equity, many often use some variation of the three-legged medical stool to gauge the performance of a health economy. We discuss and provide some historic and contemporary data for each of these components in the following sections. The discussion not only introduces the various legs of the medical stool, but also motivates and acts as a roadmap for the remaining material in this textbook.

Medical Care Costs

Although the topic of medical care costs is taken up more formally in Chapter 7, recall from our earlier discussion that medical care resources, like resources in general, are scarce at a given point in time. It follows that an opportunity cost, or a price, is associated with each and every medical care resource because of scarcity. Thus, we can think of medical care costs as representing the total opportunity costs when using various societal resources such as labor and capital to produce medical care rather than other goods and services.

Each year since 1960, actuaries at the Centers for Medicare and Medicaid Services (CMS) have collected and reported data on the uses, sources, and costs of medical care in the United States. The data can be compared across various industries in the health care sector, like hospital, physician, and nursing home services, examined in a particular year, or tracked over time. Funding sources including consumers, insurers, or government can also be examined for various types of medical care, and over time. Hence, the CMS data yield important insights with respect to how health care funds are used, where the funds come from, and how much money in total is spent on medical care in the United States.

Uses of Medical Funds

Figure 1–2 provides a percentage breakdown of the uses of health care funds in 2006. These statistics offer insight into the mix of medical goods and services actually produced and consumed in the U.S. health economy. Recall that the second basic question is “what mix of medical care ‘should be’ produced.” Also recall that more of one type of medical care means less of the others for a given size of the medical care pie.

According to the figure, 31 percent of medical care funds is spent on hospital services. The “big ticket” nature of hospital services should not be too surprising. Acutely ill individuals...
CHAPTER 1 Introduction

Physician services make up the next largest use of funds with 21 percent of the total. The dominant role of physicians makes sense because they are primary care gatekeepers and patients must often first pass through them before accessing other types of medical care, including hospitals and prescription drugs. In addition, specialty physicians, such as heart surgeons, provide important services that maintain, improve, and extend human lives. Their reimbursement reflects, in part, the value placed on remaining healthy, which is discussed in Chapter 3.

Collectively, hospital and physician services account for more than half of all health care spending, not only in 2006 but over time as well. We will learn more about the structure, conduct, and performance of the physician and hospital services markets in Chapters 12 and 13, respectively. Finally, prescription drugs (10 percent), nursing home care (6 percent), dental care (4 percent), and home health care (3 percent) represent four other major areas where medical care funds are directly spent on patient care. The prescription drug industry is taken up in Chapter 14, whereas the home health and nursing home care industries are discussed in Chapter 15.

Sources of Medical Funds

The percentage of medical funds coming directly from consumers, private insurers, and government are shown in Figure 1–3. We emphasize the word directly because all funds ultimately come from the consumer in the form of out-of-pocket payments, premiums, and/or taxes. In 2006, 54 percent of all funds spent on national health care came from the private sector, down from approximately 76 percent in 1960. The bulk of this decrease took place in the mid-1960s when two public health insurance programs—the Medicare and
Medicaid programs—were first introduced. Since 1990, the share of national health expenditure emanating directly from the private sector has dropped slightly from 59 percent.

The mix between private insurance and out-of-pocket payments has also changed in recent years. In particular, private insurance has expanded its role as a source of funds and substituted greatly for out-of-pocket payments. In 1980, for example, private health insurance provided funds for 29 percent of all health care costs in the nation and out-of-pocket payments provided another 17 percent. By 2006, slightly over one-third of national health care expenditures came from private insurers while consumers out-of-pocket payments fell to 12 percent. The greater reliance on private insurance funding reflects both a greater number of individuals and more types of medical care (e.g., pharmaceuticals and dental) covered by medical insurance. Business payments to provide health care services directly to employees, philanthropic sources, private construction, and nonpatient revenue sources (such as revenues from hospital gift shops), help to account for the remaining 8 percent of all private funds in 2006.

Figure 1–3 also shows that 46 percent of all national health spending in 2006 came from the government. Most of the government funds were spent by the Medicare and Medicaid public health insurance programs. Given that the government funds less than half of all health care spending in the nation, the United States is often looked upon as possessing a privately financed health care system. However, Woolhandler and Himmelstein (2002) offer an alternative view of the relative share of health care spending financed through private and public sources. In particular, they scrutinize the method used by CMS to measure government spending in the national health accounts and show that the government has much more responsibility than the private sector with respect to financing the U.S. health care system.
Woolhandler and Himmelstein explain that CMS includes only direct purchasing of medical care for programs such as Medicare, Medicaid, and government-owned hospitals in its measure of government spending. Consequently, public employee benefits, such as those through the Federal Employees Health Benefits Program and various state employee health insurance programs, are missing from CMS’s reported figures. Although the government supports these public health insurance programs with tax financing, private insurers administer the program on behalf of the government and are responsible for writing the actual checks. In addition, the authors point out that employer-sponsored health insurance premiums are exempted from various federal, state, and city taxes. (We take this up later in Chapter 6.) Thus, the government also implicitly helps to finance employer-sponsored health insurance through these tax preferences.

To get a better idea about the extent to which health insurance is tax financed, the authors add together the direct purchasing of medical care by government with expenditures on public employee health benefits that are tax-financed but administered by the private sector plus the value of the health insurance premium tax preference. Woolhandler and Himmelstein report that the direct spending of government equaled 45 percent of all health care costs, while public employee benefits accounted for another 5.4 percent, and the tax subsidy for health insurance premiums amounted to an additional 9.1 percent. Thus, government, at all levels, was responsible for financing nearly 60 percent of all health care costs in the United States. Thus, one might rightfully argue that, similar to other countries around the world, the government largely finances the health care system in the United States.

These estimates of Woolhandler and Himmelstein are certainly provocative. They show that tax financing represents the major source of health care funds in the United States. Indeed, tax financing accounts for an even greater share of health care costs, considering that not-for-profit health care organizations such as hospitals, behavioral health care organizations, and nursing homes are also granted preferences on income, property, and sales taxes. (We also take this up in later chapters.) How these highly credible estimates are interpreted and used in future policy discussions concerning health care reform will be interesting to see.

## Amount of Medical Care Spending

Only someone living in entire seclusion, perhaps a World War II Japanese soldier hiding somewhere on a Pacific island or someone raised in a nuclear fallout shelter of the 1950s, would be unaware of the situation involving medical care costs in the United States. Indeed, it seems that not a day goes by without a radio, television, or popular press commentator pointing, with much alarm, to the high and continually rising costs of health care. There is certainly no need to dispute those facts. According to CMS figures, the United States spent $2.1 trillion on health care or slightly over $7,000 per person in 2006. Compare that to the similar figures of $26.9 billion and $141 dollars in 1960.

These figures are potentially alarming because trade-offs may be involved. That is, the PPC tells us that high health care costs translate into lower amounts of other goods produced and consumed. Certainly, high health care costs could reflect more and better medical care, but high spending may also involve the sacrifice of other equally important

---

6. One of the authors was stationed in Guam during the Vietnam conflict. A World War II Japanese soldier was rumored to be hiding on the island. View the movie *Blast from the Past* starring Brendan Fraser to learn how growing up in a fallout shelter can affect one’s knowledge of current events.
goods and services like food, clothing, and shelter. However, the productive capacity of the U.S. health economy has changed over time—the situation may not be as bleak as the statistics show. For example, the economy may now possess more labor and capital resources and productivity-improving technologies. Thus, the PPC has likely shifted out and therefore more of one good or service can be produced without sacrificing the others.

One way of controlling for differences in the underlying productive capacity of an economy or economies is by dividing, in this case, the amount of health care spending by GDP. Greater productive capacity, resulting from higher amounts of resources and better technology, generally means a larger level of GDP and therefore more goods and services in general. With that notion in mind, Figure 1–4 shows health care spending as a percentage of GDP from 1960 to 2006.

Figure 1–4 shows that health care spending as a percentage of GDP has grown tremendously over time in the United States. Standing at 5.2 percent in 1960, that same ratio of health care spending to GDP is now about 16 percent, which means instead of spending $1 out of every $20, we now spend $1 out of every $6 on health care. However, even the rising percentage of GDP devoted to health care does not necessarily indicate other goods and services have been sacrificed. The GDP of $13 trillion in 2006 is much greater than the GDP of $526 billion in 1960. Given the health care spending to GDP ratios in the two years, spending on all other goods amounted to nearly $11 trillion in 2006 compared to $495 billion in 1960. Simply put, the greater productive capacity of the U.S. economy allowed for greater amounts of both health care and all other goods to be produced.

![Figure 1–4: National Health Care Costs as a Percentage of GDP](http://www.cms.gov)

In fact, productivity-enhancing technologies in the rest of the economy may have freed up resources for use in the health economy where the labor intensity of medical services doesn’t allow us much productivity improvement. Of course, the relative mix of goods has certainly favored the health care sector since 1960.

Figure 1–4 also shows that health care spending has not increased at the same continuous rate throughout the years. For example, health care spending grew more quickly relative to GDP prior to the 1990s. In contrast, notice that after the 1990s, the ratio of health care spending to GDP remained relatively stable during the 1993 to 1999 period. The ratio of health care costs to GDP has also remained fairly constant since 2003.7

Policy makers continue to debate the cause and desirability of rising health care costs in the United States and in other countries. Some argue that the U.S. health care system contains a lot of production inefficiency that can and should be squeezed out. Others point out that the benefits from health care more than compensate for the costs. Much of this debate is covered in various chapters of this book. It shouldn’t be too surprising that health economists are heavily involved in this debate. In fact, they often draw upon the tools that can be learned in this book when trying to make some sense of health care spending and the health care economy. The structure of a health care system certainly plays a role so that topic is taken up in Chapter 4. The material in Chapters 5, 6, 7, and 8 also add to our understanding of health care costs and how consumers, providers, insurers, markets, government, and economic incentives help to shape health care spending.

Medical Care Access

Medical care access, another leg of the medical stool, relates to the distribution question. That is: Does everyone have reasonable access to medical care on a timely basis? Timely access is often measured by the percentage of individuals with health insurance. For most people the cost of catastrophic care, such as organ transplants and cardiovascular surgery, lies beyond their financial means. But, as explained fully in Chapter 6, for a relatively small payment or premium, insurance provides access to high-cost, life-saving interventions if and when people experience severe illnesses. Thus, health insurance may be an important factor in terms of ensuring timely access to medical care. Figure 1–5 offers some information on the percentage of people without health insurance in the United States since 1940.

Before discussing the data in Figure 1–5, it should be noted that the health insurance product has changed considerably over time. Prior to the 1970s most people purchased only hospital insurance. Today people purchase health insurance for other types of medical care, as mentioned previously. Also, the amount of medical care expenditures covered by insurance has increased over the years. Thus, for the sake of consistency, it may be best to think of Figure 1–5 as showing the percentage of the U.S. population without hospital insurance.

In any case, the data in the figure show that great strides have been taken in terms of more people insured in the United States. In 1940, only 10 percent of the U.S. population possessed health insurance purchased almost entirely in the private marketplace. Even before public health insurance programs, beginning with the Medicare and Medicaid Acts

---

7 To examine the efficiency consequences of medical spending we must consider the benefits of medical goods and services in addition to their costs. That topic is taken up in Chapter 3.
of the mid-1960s, many people began purchasing private health insurance in the United States following the 1940s. By 1975 the uninsured rate in the United States dipped to about 13 percent because of both private purchases and public expansions. However, beginning around 1980, further persistent declines in the uninsured rate have not materialized. In 2006 the uninsurance rate in the United States stood at nearly 16 percent. While we take up the causes, types, and social costs of uninsurance in Chapters 6 and 11, it suffices to note that a sizeable percentage of the U.S. population lacks timely access to medical care because of their uninsured status. In addition, severe racial disparities exist with respect to uninsured status, as noted in Chapter 11. Clearly, these are two additional areas where the tools of health economies are needed to shed better light and bring about improvements in the health economy and society.

**Medical Care Quality**

The final leg of the medical stool we consider is medical care quality. As discussed more fully in Chapter 2, quality represents a complex and multidimensional concept. In keeping with the other two legs of the medical stool we confine our discussion to a single measure of quality that is easily understandable and important from a societal point of view, and for which data can be obtained over time for comparative purposes. The chosen measure is the infant mortality rate that tells us the number of children below one year of age that died as a percentage of all live births in that same year. The infant mortality rate for the United States from 1960 to 2006 is reported in Figure 1–6.
Like the uninsured rate, the infant mortality rate has improved significantly over time in the United States falling from a height of over 25 infant deaths per 1,000 live births in 1960. Although it stands to reason that rising health care spending and increased insurance coverage contributed to the decline, Chapter 2 discusses the theoretical framework and empirical findings regarding the many factors influencing health status outcomes such as infant mortality. Despite the vast improvements that have taken place over time, Figure 1–6 suggests that nearly 7 out of every 1,000 live babies in the United States do not live beyond 1 year of age. Also, the United States lags far behind when compared to other industrialized countries like Belgium, France, Italy, Japan, and the United Kingdom, which have infant mortality rates below 5 deaths per 1,000 live births. Finally, the figure does not capture the vast variations in health outcome measures, such as infant mortality, among different income, racial, and ethnic groups. Once again, the tools of health economics can prove useful for analyzing health outcomes and proposing ways of improving societal health.

### A Note on the Relation between System Structure and Performance

Many theories and empirical findings pertaining to health economics are introduced and developed in this text. Sometimes theories and empirical findings are of interest for their own sake, particularly for academicians such as the authors. But the main reason for their introduction and development is that we wish to obtain a better grasp of the operation and
performance of the real-world health economy around us. If the health economy does not perform in a socially efficient and equitable manner, then we would hope that solutions could be proposed and policies could be changed to alter that undesirable performance.

An understanding of the link between structure and performance is essential when crafting new policies. Structure plays a role in determining how people behave or conduct themselves in the health economy. Figure 1–7 shows the complex interaction between
structure and performance. A health economy is structured in a particular way, and this health economy structure is discussed in great detail in Chapter 4. Structure shows up in the ways various organizations are designed in terms of their size and scope, the mix of market activities and government involvement in the health economy, and financing and reimbursement mechanisms, among others.

This underlying structure helps to establish the prevailing incentives in a health economy and thereby influences how people, organizations, and government itself, behave. If incentives are distorted because of structural defects, then suboptimal performance likely results in terms of inefficient and inequitable outcomes. Given the suboptimal performance, solutions can be proposed and public policies can be designed to remedy the situation. In particular, policies can be changed to either indirectly affect behavior through a restructuring of the system or directly by introducing conduct remedies.

Just about every chapter in the book addresses an issue where incentives are discussed or public policy plays a role. As mentioned previously, health economists are most interested in the efficiency of outcomes because resources are scarce. Unfortunately, efficiency is often difficult to gauge or measure in practice. An alternative is to design a theoretical benchmark where efficiency can be attained; then compare the real world, in terms of the existing incentives because of its structure, to that theoretical benchmark. Our benchmark for allocative efficiency (the point at which marginal social benefit equals marginal social cost) is developed in Chapter 3. This benchmark is expanded upon in Chapter 8. The most discussion concerning public policy shows up in Chapters 9, 10, and 16.

Summary

Health economics is concerned with the determination and allocation of health resources and distribution of medical services in a society. Because resources are scarce, society must determine what amounts of medical services to produce, what kinds of medical services to produce, what mix of health care resources should be used, and who should receive the output of health care services. Answering these four basic questions involves tough trade-offs.

A health economy, like a macroeconomy in general, can be analyzed with respect to its performance. We discussed how the health economy can be assessed with regard to medical care cost, access, and quality and learned that the tools of health economics can and will be used to explore more thoroughly these components of the three-legged medical stool in subsequent chapters of this book. Controlling medical costs, access, and quality also involves trade-offs.

Finally, economic analysis can help us better understand the causes of problems relating to health and health care. The tools and concepts of health economics can also be used to find solutions and offer public policy prescriptions. The public policy prescriptions may involve structural and/or conduct remedies.

Review Questions and Problems

1. Draw a bowed-out PPC with an aggregate measure of medical services, Q, on the horizontal axis and an aggregate measure of all other goods (and services), Z, on the
vertical axis. Discuss the implications of the following changes on the quantities of medical services and all other goods.
A. A movement down along the curve.
B. A movement from the interior of the curve to a northeasterly point on the curve.
C. An increase in the quantity of labor in the economy.
D. A technological discovery that increases the production of Z.
If it were your choice, where would you choose to produce on the PPC? Why?
2. Congratulations! Upon graduating you accept a well-deserved job with XER Consulting. Your first job involves a consulting gig with a state subcommittee on health care issues. The senate health care subcommittee is considering the expansion of two existing public health programs. One program concerns additional funding for nursing homes around the state. The other program involves additional funding for community health centers around the state. In both cases the funding is supposed to be used to attract more nurses for expansion purposes. Your job involves the following four tasks:
A. Draw and use a production possibilities curve to graphically show and verbally explain to the subcommittee members the opportunity cost at a point in time of expanding any one of the programs, assuming that both of them are initially operating efficiently. Be sure to correctly label the axes and all points. Refer to the points on the graph in your explanation.
B. Use the production possibilities curve to graphically show and verbally explain how one or both programs could be expanded at a lower opportunity cost if some inefficiency or slack initially exists in the overall public health system. Refer to various points on the graph in your explanation.
C. Use the production possibilities curve to graphically show and verbally explain how both programs could be expanded at a lower opportunity cost if growth is expected for the public health care system. Refer to points on the graph in your explanation.
D. Verbally explain to the subcommittee members what factors might cause the public health care system to grow.
3. Identify the so-called three legs of the medical stool. Explain how trade-offs might take place among the three legs. If you had to choose one of the three to improve upon at the neglect of the others, which would you choose? Why?
4. Does the U.S. health care system possess a privately or publicly financed health care system? Explain.
5. What are two major uses of medical funds? How do the two major uses relate to the four basic questions?
6. At this point in the book, do you think the United States spends too much on medical care? Explain your reasoning using the PPC.
7. Explain the change in the percentage of the U.S. population with health insurance from 1940 to 1980. Can you think of any economic factors that may have caused that change? Explain the change in the percentage insured since 1980.
8. Explain the change in the infant mortality rate (IMR) in the United States since 1960. Do you think the IMR is too high in the United States? Why? What is the implication of a reduction in the IMR if we treat infant mortality rate reductions as one good on one axis of the PPC and all other goods on the other axis? What is the implication of an IMR reduction if we assume some production inefficiency initially exists in the U.S. health care system? Why?
9. In your own words, explain the general link between system structure, performance, and policy.
References


Appendix 1: Economic Models and Empirical Testing

Health economics can be considered as both a social science and a science. As a social science, the field of health economics studies people in their everyday lives and addresses issues such as obesity, alcohol abuse, and abortion. As a science, health economics offers testable hypotheses. For example, a health economist might explore empirically if people purchase more whiskey or fast food when their prices decline—the so-called law of demand. In either case, models and empirical methods are used in health economics. This appendix offers an introduction to both of these tools of health economic analysis.

Economic Models

As mentioned earlier, the PPC is an example of an economic model. Models are abstractions of reality and are used in economics to simplify a very complex world. Economic models can be stated in descriptive (verbal), graphical, or mathematical form. Usually an economic model like the PPC describes a hypothesized relation between two or more variables. For example, suppose the hypothesis is that health care expenditures, \( E \), are directly (as opposed to inversely) related to consumer income, \( Y \). That hypothesis simply means that expenditures on health care services tend to rise when consumer income increases. Mathematically, a health care expenditure function can be stated in general form as

\[
E = f(Y).
\]

Equation A1–1 implies that health care spending is a function of consumer income. In particular, health care expenditures are expected to rise with income.

An assumption underlying economic models is that all factors, other than the variables of interest, remain unchanged. For example, our hypothesis that health care expenditures are directly related to income assumes that all other likely determinants of health care spending, such as prices, tastes, and preferences, stay constant. As another example, notice in the previous analysis that the stocks of resources and technology are held constant when constructing the PPC. Indeed, economists normally qualify their hypotheses with the Latin phrase ceteris paribus, meaning “all other things held constant.” By holding other things constant, we can isolate and describe the pure relation between any two variables.

8. In fact, economics, of which health economics is a subdiscipline, touches upon history, psychology, sociology, philosophy, mathematics, and statistics.
The expenditure function in Equation A1–1 is expressed in general mathematical form, but a hypothesis or model is often stated in a specific form. For example, the following equation represents a linear expenditure function for health care services:

\[ E = a + bY, \]  

where \(a\) and \(b\) are the fixed parameters of the model. This equation simply states that health care expenditures are directly related to consumer income in a linear (rather than nonlinear) fashion. Mathematically, the parameter \(a\) reflects the amount of health care expenditures when income is zero, whereas \(b\) is the slope of the expenditure function. The slope measures the change in health care expenditures that results from a one-unit change in income, or \(\Delta E/\Delta Y\).

For example, let us assume the parameter \(a\) equals $1,000 per year and \(b\) equals one-tenth, or 0.1. The resulting health care expenditure function is thus

\[ E = 1,000 + 0.1Y. \]

Equation A1–3 implies that health care expenditures rise with income. In fact, the slope parameter of 0.1 suggests that each $1,000 increase in consumer income raises health care spending by $100.

The health care expenditure function in Equation A1–3 is represented graphically in Figure A1–1. Yearly consumer income per household is shown on the horizontal axis, and annual health care spending per household is shown on the vertical axis. According to the function, health care spending equals $3,000 when household income is $20,000 per year. Consumers earning $50,000 per year spend $6,000 per year on health care services. Note that the expenditure function clearly represents our hypothesis concerning the direct relation between income and health care spending.

Now suppose some other determinants of health care expenditures change. Although this assumption violates our implicit *ceteris paribus* condition, we can incorporate changes in other factors into the health care expenditure model fairly simply. For example, suppose people generally become sicker than before, perhaps because households have become older on average. Obviously, this change tends to increase health care spending. Assuming that the “aging” effect influences only the intercept term and not the value of the slope parameter, the expenditure function shifts upward by the yearly increase in health care spending due to the aging population. Figure A1–2 shows an example of this effect.

Yearly medical costs are assumed to increase by $500 for the typical household. Thus, the health care expenditure function shifts upward at each level of income by $500 to \(E_1\). If the aging effect also influences the percentage of additional income that people spend on health care services, the slope of the function changes as well. An increase (decrease) in the marginal propensity to spend out of income raises (lowers) the slope and rotates the expenditure function to the left (right).\(^9\)

As you can see, a model, such as this expenditure function or the PPC, is useful because it helps simplify an otherwise complex world. We can better and more easily understand

---

\(^9\) Problem 2 at the end of the chapter asks you to complete an exercise of this type.
the relation among key variables. Models are also useful because they often offer valuable insights into the necessity or relative effectiveness of various public policies. For example, we saw from the PPC that policy changes typically involve trade-offs that public policy makers should heed.

In the case of our health care expenditure function, suppose that some government agency, such as the U.S. Government Accountability Office or Congressional Budget Office, determines that $4,000 of annual household spending on health care is necessary to maintain the health of family members in the typical household. Further suppose that a study by this same government agency finds that our health care expenditure model, as reflected in Equation A1–3, represents the true relation between household income and health care spending. If so, our model suggests that households with incomes less than $30,000 tend to spend less than the necessary amount on health care. The government might use this information to determine the subsidy needed at each level of family income to reach the targeted amount of $4,000. For example, a household with $10,000 of

![FIGURE A1–1
Health Care Expenditure Function](image)

According to the expenditure function, health care spending increases with income. For example, health care spending equals $3,000 when household income equals $20,000 per year and $6,000 when household income equals $50,000 per year.
income would require a $2,000 subsidy to reach the targeted amount of health care spending whereas a household with $28,000 would need only $200.

Consequently, economic models are useful because they help simplify complex situations so we can more easily understand how things fit together. Models also are of great use for policy purposes.

Positive and Normative Analysis

Health economists perform two types of analysis. **Positive analysis** uses economic theory and empirical analysis to make statements or predictions concerning economic behavior. It seeks to answer the question “What is?” or “What happened?” For example, we might investigate the exact relation between income and health care spending. Because positive analysis provides or predictions, it tends to be free of personal values.

**Normative analysis**, on the other hand, deals with the appropriateness or desirability of an economic outcome or policy. It seeks to answer the question “What ought to be?” or “Which is better?” For example, an analyst might conclude that households with incomes less than $30,000 per year should be subsidized by the government because they are unable...
to maintain a proper level of health care spending. Naturally, this implies that the analyst is making a value judgment. Because opinions vary widely concerning the desirability of any given economic outcome and the role government should play in achieving outcomes, it is easy to see why normative statements generally spark more controversy than positive ones. For instance, when 518 health economists were asked whether the Canadian health care system is superior to the U.S. system, there was much disagreement. Fifty-two percent of the economists agreed and 38 percent disagreed with the statement. The remaining 10 percent had no opinion or lacked the information needed to respond to the question (Feldman and Morrisey, 1990).

The following sets of positive and normative economic statements should give you a better understanding of the difference between the two. Notice that the positive statements deal with what is or what will be, whereas the normative statements concern what is better or what ought to be.

**Positive:** According to Becker and Murphy (1988), a 10 percent increase in the price of cigarettes leads to a 6 percent reduction in the number of cigarettes consumed.

**Normative:** The government should increase the tax on cigarettes to prevent people from smoking.

**Positive:** A study by Hellinger (1991) estimates that the average yearly cost of treating someone with AIDS is $38,300, while the lifetime costs equal $102,000.

**Normative:** It is in our country’s best interests that the federal government take a more active role in the prevention of AIDS.

**Positive:** National health care expenditures per capita are higher in the United States than Canada.

**Normative:** To control health care expenditures, the United States should adopt a national health insurance program similar to Canada’s.

**Empirical Testing**

Empirical testing of economic theories is important for two reasons. First, economic hypotheses require empirical validation, especially when a number of competing theories exist for the same real-world occurrence. For example, some people believe medical illnesses occur randomly whereas others believe medical illness is largely a function of lifestyle. The “random” and “lifestyle” explanations represent two competing theories for medical illnesses. Empirical studies can potentially ascertain which theory does a better job of explaining illnesses.

Second, even well-accepted theories are unable to establish the magnitude of the relation between any two variables. For example, suppose we accept the theory that lifestyle is a very important determinant of health status. A question remains about the magnitude or strength of the impact lifestyle has on health status. Does a young adult who adopts a sedentary lifestyle face a 10, 20, or 50 percent chance of dying prematurely compared to an otherwise comparable individual? Empirical studies can help provide the answer to that question.

There are many different ways for researchers to conduct an empirical analysis. The method we emphasize in this book, which most economists also use, is **regression analysis**. Regression analysis is a statistical method used to isolate the cause-and-effect relation.
among variables. Our goal is to provide the reader with an elementary but sufficient understanding of regression analysis so the regression results discussed in this book can be properly interpreted. Regression analysis is explained through an example.

The example used concerns the relation between health care expenditures, \(E\), and consumer income, \(Y\). Suppose we hypothesize that health care expenditures rise with household income and want to test our theory. Health care expenditures represent the dependent variable, and income is the independent variable. Furthermore, suppose we expect a linear (or straight-line) relationship between income and health care expenditures, or

\[
(A1–4) \quad E = a + bY,
\]

where \(a\) is the constant or intercept term and \(b\) is the slope parameter. If you recall, the slope parameter in this case identifies the change in health care expenditures that results from a one-unit change in income.

Because we are interested in the actual or real-world magnitudes of the parameters \(a\) and \(b\), we will now collect a random sample of observations relating information on both medical expenditures and income. The data might be series observations on income and expenditures for a particular household over time or cross-sectional observations on income and expenditures across different households at a particular point in time. In this case, the household represents the unit of analysis but the unit of analysis could be an individual or a town, county, state, region, or country. Suppose we collect cross-sectional data on income and medical expenditures from a random survey of 30 households.

Exhibit A1–1 shows a scatter diagram illustrating our random sample of observations (only 5 of the 30 observations are illustrated for easier manageability). Notice that the scatter diagram of observations does not automatically show a linear relation between income and health care expenditures because of omitted factors that also influence spending on health care, some randomness to economic behavior, and measurement error. Our objective is to find the line that passes through those observations and provides the best explanation of the relation between \(Y\) and \(E\). One can imagine numerous lines passing through the set of observations. What we want is the line that provides the best fit to the data.

A criterion is necessary to determine which line constitutes the best fit. One popular criterion is ordinary least squares, or OLS. OLS finds the best line by minimizing the sum of the squared deviations, \(e_i\), from the actual observations and a fitted line passing through the set of observations, or

\[
(A1–5) \quad \text{Minimize } \sum e_i^2 = \sum (E_a - E_f)^2 = \sum (E_a - \hat{a} - \hat{b}Y)^2,
\]

where \(E_a\) is the actual observation on medical expenditures and \(E_f\) is fitted (or predicted) expenditures from the estimated regression line, \(\hat{a} + \hat{b}Y\). In Exhibit A1–2, we show an example of a fitted line and the resulting deviations between actual and fitted expenditures. Based upon the sample of observations, a computer program (such as SAS, SPSS, or TSP) searches for the best line using the OLS procedure. In the process of finding the best line, the intercept and slope are determined, and thus we estimate the best magnitudes for \(a\) and \(b\) that minimize the sum of the squared deviations from the actual observations.

Let’s suppose the following results are obtained from the regression analysis:

\[
(A1–6) \quad E = 2,000 + 0.2Y.
\]
EXHIBIT A1–1
Scatter Diagram of Income and Health Care Expenditures

A scatter diagram showing the actual relationship between household income and health care spending for 5 observations.

EXHIBIT A1–2
Fitted Line

The fitted line resulting from OLS and the associated deviations between the fitted and actual values.
The results would tell us that the best fitted line to the data has an intercept of $2,000 and a slope of 0.2. Although the fitted or estimated regression line provides the “best” fit compared to all other lines, we do not know yet whether it represents a “good” fit to the actual data. Fortunately, the computer estimation procedure also provides us with some goodness-of-fit information that we can use to determine if the best fit is also a reasonably good one.

The two most common and elementary goodness-of-fit measures are the coefficient of determination, $R^2$, and the $t$-statistic, $t$. The coefficient of determination identifies the fraction of the variation in the dependent variable that is explained by the independent variable. Thus, the $R^2$ ranges between 0 and 1. Researchers tend to place more faith in a regression line that explains a greater proportion of the variation in the dependent variable.

The values for the parameters $\hat{a}$ and $\hat{b}$ are average estimates rather than true values because they are based on a sample instead of all possible observations; thus, they are associated with some error. Accordingly, there will be some deviations around the average estimate for $a$ and also around the average estimate for $b$. In fact, if the deviations are very large, we cannot place much faith in the estimated value for the parameters. Indeed, the true value for $b$ may be zero. If so, no relationship exists between income and health care expenditures.

The computed $t$-statistic helps us identify how much deviation occurs around the estimated average value for the parameters of the model. A $t$-statistic of 2 or more means that the value of the estimated parameter was at least twice as large as its average deviation. A rule of thumb is that when the $t$-statistic is 2 or more, we can place about 95 percent confidence in the estimated average value for the parameter, meaning that only a 5 percent likelihood exists that the relationship could have occurred by chance. Another rule of thumb is that when the $t$-statistic is 3 or more, we can place 99 percent confidence in our estimated value for the parameter. In this case, only a 1 percent likelihood exists that the relation occurred by chance.

Regression results are generally reported similar to the following:

\[
(A1–7) \quad E = 2,000 + 0.2Y \quad R^2 = 0.47
\]

\[
(2.52) \quad (3.40) \quad N = 30
\]

The $t$-statistics are reported in parentheses below the parameter estimates. Because the $t$-statistic associated with income is greater than 3, we can place a high degree of confidence in the parameter estimate of 0.2. Also, according to the regression results, income explains about 47 percent of the variation in health care expenditures. The number of observations, $N$, is 30.

Before we move on we need to interpret the parameter estimates for Equation A1–4. The intercept term of 2,000 tells us the level of health care expenditures when income is zero. The parameter estimate of 0.2 on the income variable is much more telling and suggests that expenditures on health care will increase by 20 cents if income increases by one-dollar. If the estimated parameter was instead –0.2, it would mean that a one-dollar increase in income causes health care expenditures to decrease by 20 cents. Thus, both the sign and value of the parameter estimate convey important information to the researcher.

The regression analysis we have been discussing thus far is an example of a simple regression because there is only one independent variable. Multiple regression refers to an
analysis in which more than one independent variable is specified. For example, theory might tell us that price or tastes and preferences should also be included in an expenditure equation. The OLS procedure behind multiple regression is the same as that for simple regression and finds the best line that minimizes the squared deviations between the actual and fitted values. The computed $R^2$ identifies the variation in the dependent variable, say, health care expenditures, explained by the set of independent variables, which in our example would be price, income, and tastes and preferences. Each independent variable would be associated with an estimated parameter and $t$-statistic. For example:

$$E = 1,000 - 0.2P + 0.13Y + 0.8A \quad R^2 = 0.75$$

(2.32) (0.42) (3.23) (4.00) \(N = 30\)

where $P$ represents the price of medical services and $A$ represents the average age in the household as a proxy for tastes and preferences. According to the regression results, the independent variables collectively explain 75 percent of the variation in health care expenditures. Also, the regression results suggest that income and age both have a statistically significant direct impact on health care expenditures. Price, on the other hand, has no impact on health care expenditures according to the regression findings.

**Association versus Causation**

As mentioned previously, the intent behind multiple regression analysis is to establish a cause and effect relationship among variables. Sometimes, however, multiple regression analysis simply captures an association or correlation among variables rather than a true causal relationship. That happens most often for observational studies that involve cross-sectional or time series data but contain no correction for the circumstances behind the observed relationship. The association but lack of causation typically occurs because the underlying observations have not resulted from a randomized process with both a control and a treatment group. Figure A1–3 helps to show why an observational study may be hindered by its inability to distinguish between a causal relationship and an association.

**FIGURE A1–3**

**Association Rather Than Causation**

Office visits are associated with physical health status because an unobservable factor, such as mental health status, affects both.
The figure illustrates a simple relationship between physical health status (say a self-reported index ranging from poor to excellent physical health) and the number of physician visits (as a measure of medical care). All other measurable factors affecting physical health status, like age, gender, and income, are collapsed and captured in the variable $X$. Suppose we are investigating if more office visits help to improve, or cause, better health. However, even if the multiple regression analysis yields a statistically significant relation between the number of physician visits and more favorable health we cannot be certain if the evidence supports a causal relationship. The uncertainty holds for two reasons.

First, a third unobservable and therefore immeasurable factor, $Z$, that cannot be included in $X$, may simultaneously affect both the number of physician visits and physical health status and thereby produce the observed association. For example, suppose we cannot properly and completely measure mental health status (e.g., the severity of depression) and mental health status influences both the self-reported physical health index and the likelihood of visiting a physician. Perhaps, severely depressed individuals simultaneously downgrade their physical health status and become more reclusive so they fail to visit their physician. If so, any observed correlation between physician visits and physical health status, in the presence of this important omitted unobservable variable, may not reflect causation.

Second, reverse causality may pose a problem when attempting to draw inferences about the direction of causal relationships from regression results. That is, physical health status, the dependent variable in Figure A1–3, may influence the number of physician visits, the independent variable. For example, state governments may pursue policies to encourage more doctors per person in areas with the highest infant mortality rates. Or, pregnant mothers may be more likely to seek out physicians when they suspect the health of their infants may be at greater risk. Hence, the regression results from an observational study would actually reflect a reverse effect—health status causes visits.

As a result, investigators often use various methods to identify or isolate causal relationships. Basically, some type of identification strategy is necessary to distinguish a causal relationship from an association. One strategy randomly assigns people or households to different situations or categories and conducts a controlled behavioral experiment. Following our same example, on a random basis, various individuals might be required to visit the doctor a certain number of times per year. Some individuals may not be allowed any physician visits at all, and others may be forced to visit their doctor ranging from one to ten times per year, regardless of their income, observable mental health status, or other personal characteristics. The random assignment of households corrects for any self-selection bias that results when individuals with different (unobservable) mental health states are allowed to choose the number of doctor visits.

The analyst then studies the relation between the number of office visits and physical health status, while controlling for other observable measures that may also affect health status using a technique such as multiple regression analysis. The hypothesis is that physical health status improves with more office visits—ceteris paribus. As you might expect, randomized social experiments of this kind offer valuable insights but are very expensive to conduct. In addition, the health of some individuals might be seriously compromised if they are not permitted to visit the doctor a reasonable number of times per year. Hence large social experiments are rarely conducted. In Chapter 5, we will discuss the RAND
Health Insurance Study of the 1970s which randomly assigned households to different health plans and investigated various hypotheses relating to health and health care.

A **natural experiment**, an alternative identification strategy, arises when some type of external global policy, unrelated to other determinants of physical health status, produces an uncontrollable shock in the medical care received by a treatment group. Changes in the health outcomes of this treatment group are then compared to health outcomes of the control group that did not experience that same external shock but otherwise faced fairly similar circumstances. The uncontrollable nature of the policy shock prevents self-selection.

For example, suppose the government sharply cuts funding for various public health insurance plans such that some low-income people are randomly terminated from the programs. Those individuals terminated from the programs represent the treatment group and those continuing in the programs represent the control group. After a given period, we then gather data on physical health status and other determinants of health status including age, gender, and income.

In the multiple regression analysis, physical health status serves as the dependent variable. The independent variables include a 0 or 1 dummy variable identifying if the individual was subjected to the policy shock or not, and other measurable determinants of physical health status. Assuming 1 represents an individual in the treatment group, we would expect a negative coefficient on the dummy variable because termination from the programs causes poorer health, all other factors held constant.

Several natural experiments have studied the effect of medical care program terminations (such as veteran or maternal health benefits) on the health outcomes of a treatment group compared to an otherwise similar control group for which the termination did not occur (Levy and Meltzer, 2001). While this method offers a valuable way of identifying the existence of causal relationships, various drawbacks exist. First of all, not many policy shocks occur in practice for testing various hypotheses. Even when they do, the so-called treatment and control groups may not be randomly selected. For example, in some of the studies just cited, only those individuals with less severe illnesses were terminated from the medical care programs.

The third identification strategy is called the **instrumental variables approach**. To conduct the instrumental variables approach, in the context of our example, a variable (i.e., an instrument) or a set of variables that affect the number of office visits but not physical health status must be found. For instance, the distance of each household from the physician’s office might be used as an instrument because it could be argued that distance helps to determine the number of office visits (i.e., convenience), but not physical health status.

If so, a multiple regression technique called two stage least squares can be employed to examine the extent to which distance affects the number of physician visits in the first stage of the estimation procedure and then the effect of physician visits on physical health status in the second stage. This technique essentially purges some of the association between physician visits and physical health status resulting from the third variable problem or reverse causality. That is, we can identify any change in physical health that results from a change in the number of office visits because of less or greater convenience.

The instrumental variables approach is one of the more popular methods for identifying causal relationships. However, in practice, it is often difficult to find a suitable set of
instruments. This is particularly true for health economic analyses where many variables are highly correlated with one another such as the consumption of medical care, health insurance status, income, and health status—it is very hard to find a factor of set of factors that affect one but not the others.

The final method to identify a causal relationship is referred to as the **fixed effects model**. A panel data set, which combines both cross-sectional and time series data, is necessary to use a fixed effects model. The same 100,000 people over 10 years or 50 states over 20 years represent examples of panel data sets. Because of the time dimension, we can track how the same cross-section of observations reacts to changes in various factors over time. More importantly, a 0/1 dummy variable for each cross-section observation in the sample can be specified in the multiple regression equation to control for unobservable heterogeneity (i.e., unobservable differences among the cross-section observations).

Recall from our running example that we are unable to control for the severity of mental depression and that omitted variable creates a third variable problem. Assuming each individual's state of mental depression is fairly constant over time, the set of cross-section dummy variables or fixed effects essentially helps to control for mental health status differences as well as any other unobservable differences among the individuals in the sample. This reduces the likelihood of a third variable problem and allows the researcher to better identify a causal relationship.

For that reason, most of the statistical research today in health economics involves a fixed effects model. There are a couple of shortcomings associated with the fixed effects approach, however. First, data requirements are much greater. Data for the same cross-section of observations must be obtained and inputted for a number of years. But with greater amounts of data available on-line and in predetermined formats, that shortcoming is becoming less troublesome. Second, the fixed effects model assumes that the unobservable heterogeneity, e.g., severity of mental depression, is relatively constant over time. If the unobservable variable changes over time, then the third variable problem may not be eliminated and the empirical results may reflect an association instead of a causal relationship. When a social or natural experiment cannot be performed, a preferred identification strategy combines an instrumental variables approach along with a fixed effects model.

**Summary**

Economic models and empirical testing of hypotheses are important for making sense of the real world, for advancing knowledge, and for public policy purposes. Economic models help to organize our thoughts about the relationship among key variables by helping to simplify an otherwise complex world. Positive analysis cannot be performed without economic models and normative analysis should be based on solid positive theory.

Empirical evidence should also be based on sound economic theory. That is, the variables specified in a multiple regression equation should be based on economic reasoning rather than ad hoc notions. Knowing the quantitative magnitude of the relationships among variables provides important insights into the relative effectiveness of various policies. As a result, choosing the best policy often requires hard empirical evidence.

We recognize that learning the material in this appendix does not make the reader an econometrician. Econometrics is way too complex for that to happen. The material does,
however, introduce the reader to the general idea behind the empirical testing of health economic hypotheses. It also exposes the reader to some of the pitfalls involved and several techniques for dealing with these pitfalls. The basic idea is that all multiple regression models are not created equally; some are clearly better than others. We invite you to learn more about the theory and practice of econometrics.¹⁰

Review Questions and Problems

1. Determine whether the following statements are based on positive or normative analysis. Be sure to substantiate your answers.
   A. Prices of physician services should be controlled by the government because many citizens cannot afford to pay for a visit to a physician.
   B. According to Tosteson et al. (1990), a 25 percent drop in the number of people who smoked in 1990 would reduce the incidence of coronary heart diseases by 0.7 percent by the year 2015.
   C. Rising health care costs have forced numerous rural hospitals to close their doors in recent years.
   D. According to government statistics, in 1989 7.2 deaths per 100,000 residents were alcohol induced. To decrease this number, the government should impose higher taxes on alcohol.

2. Suppose a health expenditure function is specified in the following manner:

   \[ E = 500 + 0.2Y, \]

   where \( E \) represents annual health care expenditures per capita and \( Y \) stands for income per capita.
   A. Using the slope of the health expenditure function, predict the change in per capita health care expenditures that would result from a $1,000 increase in per capita income.
   B. Compute the level of per capita health care spending when per capita income takes on the following dollar values: 0; 1,000; 2,000; 4,000; and 6,000.
   C. Using the resulting values for per capita health care spending in part B, graph the associated health care expenditure function.
   D. Assume that the fixed amount of health care spending decreases to $250. Graph the new and original health care functions on the same graph. What is the relation between the original and new health care expenditure functions?
   E. Now assume that the fixed amount of health care spending remains at $500 but the slope parameter on income decreases to 0.1. Graph both the original and new health care expenditure functions. Explain the relation between the two lines.

¹⁰The website for the text at http://www.cengage.com/economics/santerre contains another more formal econometric appendix written by Bruce Carpenter of Mansfield University. It goes into great detail on the specifics behind multiple regression analysis, logarithmic functions, and how elasticities can be determined with the estimated coefficients among other topics. Studenmund (2006) offers a good introduction to econometric issues. Also, Dowd and Town (2002) offer a worthwhile discussion of causation versus association.
3. Victor Fuchs (1996) lists the following questions in an article in *The Wall Street Journal*. Identify whether the following questions involve positive or normative analysis. All the questions deal with a Republican plan to reform Medicare, the public health insurance program for the elderly.

A. How many Medicare beneficiaries will switch to managed care?
B. How much should the younger generation be taxed to pay for the elderly?
C. Should seniors who use less care benefit financially, or should they subsidize those who use more care?
D. How many Medicare beneficiaries will switch to medical savings accounts (see Chapter 16)?
E. What effect will these changes have on utilization?
F. How much should society devote to medical interventions that would add one year of life expectancy for men and women who have already passed the biblical “three score and ten”?
G. Will senior citizens’ choices about types of coverage depend on their health status?
H. If the rate of spending growth is reduced to 6 percent from 10 percent a year, what will happen to the growth of medical services? To physician incomes?

4. Indentify two purposes of empirical testing.

5. Suppose you are explaining the technique behind OLS to a statistically-challenged but otherwise intelligent uncle of yours. Further suppose the statistical relationship concerns one between the number of physician visits and physical health status. Don’t worry about drawing causality but only explaining the OLS technique itself. Explain to him how OLS fits a line to a set of observations. You might want to use a scatter diagram and an equation for a line to make your point.

6. Suppose you are presented with the following regression equation involving health care expenditures and its determinants, where all of the variables have been defined previously.

\[
E = 500 - 25P + 0.20Y - 1.2A
\]

\[
R^2 = 0.30
\]

\[
(1.21) (2.45) (0.43) (4.13)
\]

\[N = 1,000\]

a. What percent of the variation in health care spending is explained by the various independent variables?
b. Which of the independent variable possess a statistical significant impact on health care spending? What do the results suggest about the relation between income and health care spending?
c. Supposing that both \(P\) and \(E\) are measured in dollars, interpret the coefficient estimate on \(P\).
d. What does the coefficient estimate on \(A\) suggest about the relation between age and health care spending?
e. Can you think of any omitted variables that might cause our estimates to be suspect?

7. Some years ago several researchers found a correlation between cigarette smoking and suicides. Do you think this correlation reflects an association or a causal relationship? Why? If it reflects an association, can you think of a plausible third variable?

8. What are meant by the third variable problem and reverse causation?
9. In your own words, explain the difference between a social experiment and a natural experiment.
10. In your own words, explain how the instrumental variables and fixed effects approaches deals with the third variable problem.

References


The disintegration of the Soviet Union, which many Americans viewed on their televisions with the collapse of the Berlin Wall on November 6, 1989, emerged as a major turning point in the twentieth century and radically changed the lives of millions of people. As a case in point, in just five short years, from 1989 to 1994, the life expectancy of men in Russia fell by 6.6 years. For women over the same time period, life expectancy fell by 3.3 years. Brainerd and Cutler (2005) investigate the impact of five major trends on the increase in mortality rates in Russia: the deterioration of the health care system, the increase in traditional risk factors for cardiovascular disease such as smoking, the increase in alcohol consumption, changes in diet, and material deprivation. Overall, they find that about half of the increase in mortality in Russia was brought about by increased alcohol consumption and the stress that accompanied the transition to a market economy. The other three major trends did not appear to statistically impact the increase in mortality rates.

The study by Brainerd and Cutler illustrates the important roles that medical care, lifestyle, socioeconomic conditions, and the environment play in the overall health of the people in a country. This chapter explores these relationships by establishing the theoretical and empirical connection between health and various factors such as medical care. In particular, this chapter:

- discusses the concepts of health and medical care
- introduces utility analysis to explain why people desire health
- utilizes production theory to explain the making of health
- reviews the empirical results concerning the factors that influence health
- discusses the historical impact of public health on health outcomes.
What Is Health?

The *Mosby Medical Encyclopedia* (1992, p. 360) defines health as “a state of physical, mental, and social well-being and the absence of disease or other abnormal condition.” Economists take a radically different approach. They view health as a durable good, or type of capital, that provides services. The flow of services produced from the stock of health “capital” is consumed continuously over an individual’s lifetime (see Grossman, 1972a, 1972b). Each person is assumed to be endowed with a given stock of health at the beginning of a period, such as a year. Over the period, the stock of health depreciates with age and may be augmented by investments in medical services. Death occurs when an individual’s stock of health falls below a critical minimum level.

Naturally, the initial stock of health, along with the rate of depreciation, varies from individual to individual and depends on many factors, some of which are uncontrollable. For example, a person has no control over the initial stock of health allocated at birth, and a child with a congenital heart problem begins life with a below-average stock of health. However, we learn later that medical services may compensate for many deficiencies, at least to some degree. The rate at which health depreciates also depends on many factors, such as the individual’s age, physical makeup, lifestyle, environmental factors, and the amount of medical care consumed. For example, the rate at which health depreciates in a person diagnosed with high blood pressure is likely to depend on the amount of medical care consumed (is this person under a doctor’s care?), environmental factors (does he or she have a stressful occupation?), and lifestyle (does the person smoke or have a weight problem?). All these factors interact to determine the person’s stock of health at any point in time, along with the pace at which it depreciates.

Regardless of how you define it, health is a nebulous concept that defies precise measurement. In terms of measurement, health depends as much on the quantity of life (that is, number of life-years remaining) as it does on the quality of life. Quality of life has become an increasingly important issue in recent years due to the life-sustaining capabilities of today’s medical technology. The issue gained national prominence in 1976 when, in a landmark court decision, the parents of Karen Ann Quinlan were given the right to remove their daughter, who was in a persistent vegetative state, from a ventilator. Because the quality of life is a relative concept that is open to wide interpretation, researchers have wrestled with developing an instrument that accurately measures health. In Chapter 3, we will discuss some of these measures.

Why Good Health? Utility Analysis

As mentioned earlier, health, like any other durable good, generates a flow of services. These services yield satisfaction, or what economists call utility. Your television set is another example of a durable good that generates a flow of services. It is the many hours of programming, or viewing services, your television provides that yield utility, not the set itself.

As a good, health is desired for consumption and investment purposes. From a consumption perspective, an individual desires to remain healthy because she or he receives utility from an overall improvement in quality of life. In simple terms, a healthy person feels great and thus is in a better position to enjoy life. The investment element concerns the relation between health and time. If you are in a positive state of health, you allocate less time to sickness and therefore have more healthy days available in the future to work and enhance your income or to pursue other activities, such as leisure. Economists look at
education from the same perspective. Much as a person invests in education to enhance the potential to command a higher wage, a person invests in health to increase the likelihood of having more healthy days to work and generate income.

The investment element of health can be used to explain some of the lifestyle choices people make. A person who puts a high value on future events is more inclined to pursue a healthy lifestyle to increase the likelihood of enjoying more healthy days than a person who puts a low value on future events. A preference for the future explains why a middle-aged adult with high cholesterol orders a salad with dressing on the side instead of a steak served with a baked potato smothered in sour cream. In this situation, the utility generated by increasing the likelihood of having more healthy days in the future outweighs the utility received from consuming the steak dinner. In contrast, a person who puts a much lower value on future events and prefers immediate gratification may elect to order the steak dinner and ignore the potential ill effects of high cholesterol and fatty foods.

Naturally, each individual chooses to consume that combination of goods and services, including the services produced from the stock of health, which provides the most utility. The isolated relation between an individual’s stock of health and utility is captured in Figure 2–1,
FIGURE 2–2
The Marginal Utility Curve for Health

The MU curve illustrates the relation between marginal utility and the stock of health, and it is downward sloping because of the law of diminishing marginal utility. The shape of curve reflects the notion that each additional improvement in health results in a smaller increase in utility than the previous one.

where the quantity of health, $H$, is measured on the horizontal axis and the level of utility, $U$, is represented on the vertical axis.\(^1\) The positive slope of the curve indicates that an increase in a person’s stock of health directly enhances total utility. The shape of the curve is particularly important because it illustrates the fundamental economic principle of the law of diminishing marginal utility. This law states that each successive incremental improvement in health generates smaller and smaller additions to total utility; in other words, utility increases at a decreasing rate with respect to health.

For example, in Figure 2–1 an increase in health from $H_0$ to $H_1$ causes utility to increase from $U_0$ to $U_1$, while an equal increase in health from $H_2$ to $H_3$ generates a much smaller increase in utility, from $U_2$ to $U_3$. In the second case, the increase in utility is less when the stock of health is greater because of the law of diminishing marginal utility. The implication is that a person values a marginal improvement in health more when sick (that is, when having a lower level of health) than when healthy. This does not mean every individual derives the same level of utility from a given stock of health. It is possible for two or more people to receive a different amount of utility from the same stock of health. The law

\(^1\) To simplify matters, we ignore the intermediate step between the health stock, the services it provides, and the utility received from these services and assume that the stock of health directly yields utility.
of diminishing marginal utility requires only that the addition to total utility decreases with successive increases in health for a given individual.

Another way to illustrate the law of diminishing marginal utility is to focus on the marginal utility associated with each unit of health. Marginal utility equals the addition to total utility generated by each successive unit of health. In mathematical terms,

\[
MU_H = \frac{\Delta U}{\Delta H},
\]

where \(MU_H\) equals the marginal utility of the last unit of health consumed and \(\Delta\) represents the change in utility or health. In Figure 2–1, Equation 2–1 represents the slope of a tangent line at each point on the total utility curve. The bowed shape of the total utility curve implies that the slope of the tangent line falls as we move along the curve, or that \(MU_H\) falls as health increases.

Figure 2–2 captures the relation between marginal utility and the stock of health. The downward slope of the curve indicates the law of diminishing marginal utility holds because each new unit of health generates less additional utility than the previous one.

What Is Medical Care?

Medical care is composed of myriad goods and services that maintain, improve, or restore a person’s health. For example, a young man might have shoulder surgery to repair a torn rotator cuff so that he can return to work, an elderly woman may have hip replacement surgery so she can walk without pain, or a parent may bring a child to the hygienist for an annual teeth cleaning to prevent future dental problems. Prescription drugs, wheelchairs, and dentures are examples of medical goods, while surgeries, annual physical exams, and visits to physical therapists are examples of medical services.

Because of the heterogeneous nature of medical care, units of medical care are difficult to measure precisely. Units of medical care are also hard to quantify because most represent services rather than tangible products. As a service, medical care exhibits the four \(I\)s that distinguish it from a good: intangibility, inseparability, inventory, and inconsistency (Berkowitz et al., 1989).

The first characteristic, \textit{intangibility}, means that a medical service is incapable of being assessed by the five senses. Unlike a new car, a steak dinner, or a new CD, the consumer cannot see, smell, taste, feel, or hear a medical service.

\textit{Inseparability} means that the production and consumption of a medical service take place simultaneously. For example, when you visit your dentist for a checkup, you are consuming dental services at the exact time the dentist is producing them. In addition, a patient often acts as both producer and consumer. Without the patient’s active participation, the medical product is likely to be poorly produced.\(^2\)

\textit{Inventory} is directly related to inseparability. Because the production and consumption of a medical service occur simultaneously, health care providers are unable to stockpile or maintain an inventory of medical services. For example, a dentist cannot maintain an inventory of dental checkups to meet demand during peak periods.

---

\(^2\) Educational services, like medical services, require the consumer’s active participation; that is, education is likely to be poorly provided when the student plays a passive role in the process.
Finally, **inconsistency** means that the composition and quality of medical services consumed vary widely across medical events. Although everyone visits a physician at some time or another, not every visit to a physician is for the same reason. One person may go for a routine physical, while another may go because he needs heart bypass surgery. The composition of medical care provided or the intensity at which it is consumed can differ greatly among individuals and at different points in time.

The quality of medical care is also difficult to measure. Quality differences are reflected in the structure, process, and/or outcome of a medical care provider (Donabedian, 1980, 1988). **Structural quality** is reflected in the physical and human resources of the medical care provider, such as the facilities (level of amenities), medical equipment (type and age), personnel (training and experience), and administration (organization structure). **Process quality** reflects the specific actions health care providers take on behalf of patients in delivering and following through with care. Process quality might include access (waiting time), data collection (background history and testing), communication with the patient, and diagnosis and treatment (type and appropriateness). **Outcome quality** refers to the impact of care on the patient’s health and welfare as measured by patient satisfaction, work time lost to disability, or postcare mortality rate. Because it is extremely difficult to keep all three aspects of quality constant for every medical event, the quality of medical services, unlike that of physical goods, is likely to be inconsistent.

As you can see, medical care services are extremely difficult to quantify. In most instances, researchers measure medical care in terms of either availability or use. If medical care is measured on an availability basis, such measures include the number of physicians or hospital beds available per 1,000 people. If medical care is measured in terms of use, the analyst employs data indicating how often a medical service is actually delivered. For example, the quantity of office visits or surgeries per capita is often used to represent the amount of physician services rendered, whereas the number of inpatient days is frequently used to measure the amount of hospital or nursing home services consumed.

### The Production of Good Health

Health economists take the view that the creation and maintenance of health involves a production process. Much as a firm uses various inputs, such as capital and labor, to manufacture a product, an individual uses medical inputs and other factors, such as a healthy lifestyle, to produce health. The relation between medical inputs and output can be captured in what economists call a production function. A **health production function** indicates the maximum amount of health that an individual can generate from a specific set of inputs in a given period of time. In mathematical terms it shows how the level of output (in this case, health) depends on the quantities of various inputs, such as medical care. A generalized short-run health production function for an individual takes the following form:

\[
\text{Health} = H(\text{medical care, technology, profile, lifestyle, socioeconomic status, environment})
\]

where **health** reflects the level of health at a point in time; **medical care** equals the quantity of medical care consumed; **technology** refers to the state of medical technology at a given point in time; **profile** captures the individual’s mental and physical profile as of a point
3. However, we should not rule out the possibility that poor health status or an illness might be created by additional medical services. An illness created by a medical care encounter is referred to as an iatrogenic disorder, “a condition caused by medical personnel or procedures or through exposure to the environment of a health-care facility” (Mosby Medical Encyclopedia, p. 401). For example, a physician may accidentally harm a patient by prescribing the wrong medicine for a given medical condition.
individual makes an initial visit and several follow-up visits to a physician’s office for a specific illness or treatment over a given period of time. It is very likely that the first few visits have a more beneficial impact on the individual’s stock of health than the later visits. Thus, each successive visit generates a smaller improvement in health than the previous one.

The relation between health and medical care can also be viewed from a marginal perspective, where the marginal product of medical care represents the incremental improvement in health brought about by each successive unit of medical care consumed, or

\[ MP_q = \frac{\Delta H}{\Delta q}, \]

where \( MP_q \) equals the marginal product of the last unit of medical care services consumed. The law of diminishing marginal productivity holds that the marginal product of medical care diminishes as the individual acquires more medical care. A graph of this relationship appears as a negatively sloped curve in Figure 2–4.4

The other variables in the health production function can also be incorporated into the analysis. In general terms, a change in any one of the other variables in the production function alters the position of the total product curve. The total product curve may shift in some

---

4. As in utility analysis, the marginal product of medical care equals the slope of a tangent line drawn to every point on Figure 2–3.
instances and/or rotate in others. In the latter case, the curve rotates because the marginal productivity of medical care has changed in response to the change in the other factors.

New medical technologies have profoundly affected all aspects of the production of medical care. In the broadest of terms, examples of new technologies include the development of sophisticated medical devices, the introduction of new drugs, the application of innovative medical and surgical procedures, and most recently, the use of computer-supported information systems, just to name a few. According to Cutler and Huckman (2003) and Cutler and McClellan (2001), technological change can result in treatment expansion, treatment substitution, or some elements of both. Treatment expansion occurs when more patients are treated by a new medical intervention, perhaps because of a higher success rate or lower risks to health. Treatment substitution occurs when the new technology substitutes for or replaces an older one.

In the context of our health production model, the development and application of a new medical technology causes the total product curve to pivot and rotate upward because the marginal productivity of each unit of medical care consumed increases, as illustrated in Figure 2–5. Notice that the total product curve rotates upward from TP0 to TP1 and each unit of medical care consumed now generates a greater amount of health. The movement

**FIGURE 2–5**

The Effect of Technological Change on the Total Product Curve for Medical Care

The total product curve shifts upward with the development and application of new medical technology because of an increase in the marginal product of medical care. A movement from point A to point B illustrates the case in which a new technology results in a simultaneous increase in the amount of medical care consumed and improvement in health. A movement from point A to point C depicts the case in which the new medical technology has no impact on health but results in less consumption of medical care.
from point A to point B in Figure 2–5 illustrates the case in which the improvement in medical technology brings about an increase in the amount of medical care consumed from $q_0$ to $q_1$ along with an improvement in health from $H_0$ to $H_1$. This movement represents the treatment expansion resulting from the new medical technology. Movement from point A to point C illustrates the situation in which the new technology has no impact on health but results in less consumption of medical care from $q_0$ to $q_2$. In this case, the new technology is cost saving, everything else held constant. It should be noted that the increase in the marginal product of medical care brought about by the medical technology also causes the marginal product curve to shift to the right.

The profile variable in Equation 2–2 depends on a host of variables and controls for such items as the person’s genetic makeup, mental state, age, gender, and race/ethnicity as of a given point in time (such as the beginning of the year). Any change in the profile variable affects both the intercept term and the slope of the health production function. For example, an individual’s genetic makeup may make him or her a candidate for prostate or breast cancer. If this individual gets cancer for that reason, then his or her total product curve shifts downward. That is because overall health has decreased regardless of the amount of medical care consumed. The total product curve is also likely to rotate downward at the same time because the marginal product of medical care should decrease as the profile worsens. The total product curve rotates downward because an otherwise healthy person is likely to respond more favorably to medical treatments for a given medical complication than one who is less healthy. Both of these changes are represented in Figure 2–6, where the total product curve shifts and rotates downward at the same time from $TP_0$ to $TP_1$. The marginal product curve for medical services also shifts to the left, because each incremental unit of medical care now brings about a smaller improvement in health.

The effect of age on the production of health is relatively straightforward. Age affects health through the profile variable. As an individual ages and deteriorates physically, both health and the marginal product of medical care are likely to fall. In addition, the rate at which health depreciates over the period is also likely to increase with age. This causes the total product curve to shift downward and flatten out. The decrease in the marginal product of medical care also causes the marginal product curve to shift to the left.\(^5\)

Lifestyle variables consider the impact of personal health habits on the production of health. Personal habits include such things as whether the person smokes, drinks excessively, leads a sedentary lifestyle, is overweight, or has an improper diet. For example, consider a newly health-conscious individual who decides that a change in lifestyle is in order. After a regimen of diet and exercise, this person loses some weight and improves his or her physical conditioning. As a result of this change in lifestyle, the individual’s level of health and the marginal product of medical care should increase. This causes the total product curve to shift and rotate upward.

As is the case with improvements in personal habits, improved socioeconomic conditions cause the intercept term and the marginal product of medical care to increase. For example, since education is likely to make the individual a more efficient producer of health independently of the amount of medical care consumed, the total product curve shifts upward. An individual with more education is likely to better understand the positive

\(^5\) The impact of gender on the total and marginal product curves is left to the reader and is the focus of a review question at the end of this chapter.
impact of a healthy diet on health. The total product curve also steepens, or the marginal product of medical care increases, because education allows the person to utilize each unit of medical care consumed more effectively. For example, an educated individual may be more inclined to understand and follow a physician’s advice concerning diet and exercise after undergoing a heart bypass operation. In addition, she or he may be able to recognize a medical problem early and seek medical care quickly when the effectiveness of medical treatment is generally at its maximum.

However, we cannot rule out the reverse effect that health influences education, particularly during childhood. Take the case of a child with chronic asthma where an asthma attack can be brought on by any number of events such as exposure to allergies or viral infections, and physical exertion. As a result, a child with chronic asthma is more likely to miss school, learn less while attending school, and in the end acquire less education. Over time, what the researcher may observe is a less healthy adult with only a modest level of education.
Some analysts have hypothesized that the relation between education and health is far more complex. For example, Fuchs (1979) argues that the acquisition of education and health depends on the value people place on future events, or the rate at which they discount future events. Individuals who place a high value on future benefits and are willing to postpone gratification are inclined to acquire more education and pursue a healthier lifestyle when they are young. This is because they want to reap the rewards of a higher income and a longer life that more education and a healthier lifestyle can bring. On the other hand, individuals who place a low value on future events and desire immediate gratification are not likely to acquire significant amounts of education or to follow a healthy lifestyle because they have adopted a “live for today” attitude. Thus, according to Fuchs, higher levels of education may be associated with better health not because there is a direct link between the two variables but because both variables are directly correlated with a third factor, the degree to which future events are valued.

The impact of income on health is also complex and is referred to as the “income gradient” in the literature “to emphasize the gradual relationship between the two: health improves with income throughout the income distribution” (Deaton, 2002, p. 14). Income is likely to indirectly impact health through a number of pathways. An increase in income provides the individual the means to consume more medical care. In addition, a more affluent individual is likely to be more educated, pursue a healthier lifestyle, and live in a safer environment, all of which contribute to improved health. For example, a more affluent individual may live in a suburban community where the crime rate is low, access to drugs and alcohol is limited, and quality medical care is available just around the corner. Income may also have a direct impact on health, although the net effect is far from clear. On the one hand, a wealthier individual may be employed in a safer work environment where the risk of a work-related accident or illness is slim. On the other hand, a wealthier individual may be employed in a more stressful occupation, which can adversely impact health.

In recent years an extensive body of literature has developed that examines whether the distribution of income impacts health, and the income-health hypothesis has taken on a variety of forms. According to the literature (Lynch et al., 2004; Wagstaff and van Doorslaer, 2000), the various hypotheses that have been offered over time can be classified into four broad categories: the absolute income hypothesis, the relative income or deprivation hypothesis, the relative position hypothesis, and the income inequality hypotheses.

The absolute income hypothesis simply states that an individual’s absolute income is positively related to health for the reasons discussed previously. The relative income or deprivation hypothesis posits that an individual’s income relative to some social group average impacts overall health. Put in more definable terms, it is a person’s income relative to some critical level such as the poverty line in the United States that matters. The presumption is that anyone with an income below the poverty line lacks the ability to acquire the basic necessities, such as health care.

The relative position hypothesis emphasizes that one’s social position in the income distribution also impacts health. For example, those at the bottom of the income scale in the United States may become frustrated and feel left behind by the “American dream” despite the fact that they have enough income to live in reasonable housing and receive adequate health care. Out of a sense of discouragement, these people may tend to give up and pursue a lifestyle detrimental to their health that could involve increased alcohol consumption, smoking, and obesity.
Finally, the income inequality hypothesis states that the distribution of income itself directly impacts health. For example, greater income inequality may create an incentive for government to limit spending on social programs that have a direct bearing on health in an attempt to lower taxes. Greater income inequality may also lead to an erosion of social capital, defined as “those features of social organizations—such as the extent of interpersonal trust between citizens, norms of reciprocity, and vibrancy of civic organizations—that facilitate cooperation for mutual benefit” (Kawachi and Kennedy, 1999, p. 221). As a result, the poor may find their public health needs largely ignored by society at large.

An adjustment in a person’s physical environment is also likely to affect the total product curve. For example, an individual with an asthmatic condition might move from Los Angeles, where smog is intense, to a community on the far outskirts of the city. Or the person’s spouse may give up smoking to decrease the level of secondhand smoke in the home. As a result, the probability that this person will succumb to a respiratory ailment diminishes. Both of these changes cause the total product curve to shift and rotate upward.

In short, health production theory suggests that a variety of factors, such as the individual’s profile, medical care, state of medical technology, lifestyle, socioeconomic status, and environment, interact to determine health. The theory also suggests that health increases at a diminishing rate with respect to greater amounts of medical care consumed, provided all other inputs remain constant. If any other inputs in the production process change, the impact of medical care on health is also likely to change. The effect of any one nonmedical input on health is also likely to exhibit diminishing returns—all other inputs held constant. For example, running two miles a day may reduce someone’s weight by 15 pounds over a six-month period. It is doubtful, however, that an additional two miles per day of running could produce additional 15 pounds of weight loss during the next six-month period.

Before we conclude this section, you should be aware that recently Jacobson (2000), Bolin et al. (2002), and others have extended the Grossman model and developed a number of sophisticated mathematical models that focus on the family rather than the individual as the main producer of health. While these models are beyond the scope of this book, they represent a valuable addition to the literature. The common theme is that individual decisions to invest in health are made within the context of a family and that any decision on the part of one family member regarding investments in health impacts the health investment decisions of others in the family. For example, a learning-disabled child may provide an incentive to a mother to invest more in her own health to ensure that she will have the time to aid her child. These theoretical developments provide a number of challenges to researchers as they strive to understand the complex relationships between family members and individual health-related decisions.

Empirical Evidence on the Production of Health in the United States

Health economists have long been trying to understand the complex nature in which medical care and other factors interact to improve, maintain, and restore health. That quest has led researchers to develop a variety of sophisticated estimation models that
find their theoretical underpinnings in Equation 2–2 to empirically examine the production of health. Using the literature as our guide, we review the empirical evidence concerning the characteristics associated with the production of health for adults and infants.

The Determinants of Health among Nonelderly Adults

Medical Care and Health  To no one’s surprise, the literature has found the consumption of medical care has a positive impact on the production of adult health. However, the results also indicate that quantitatively, the impact is relatively small. For example, Hadley (1982) finds that a 10 percent increase in per capita medical care expenditures results in only a 1.5 percent decrease in the adult mortality rate. His result confirms those of an earlier study conducted by Auster et al. (1969), who estimate that a 10 percent increase in medical services leads to a 1 percent drop in the age-adjusted mortality rate. Sickles and Yazbeck (1998) find that a 10 percent increase in health-related consumption leads to about a 0.3 percent improvement in health as measured by a comprehensive health index that considers a number of quality-of-life variables. Finally, based upon a random assignment of households to different health plans, Newhouse et al. (1993) find that households in low coinsurance plans received more medical care yet possessed virtually the same level of health as those households in high coinsurance plans, ceteris paribus. Enthoven (1980) has referred to the small marginal impact of medical care services on the health status of adults as “flat-of-the-curve” medicine. In the context of Figure 2–3, this means the typical adult consumes medical services at the point where the slope of the total product curve or marginal product of medicine is near zero.6

If, as the empirical evidence indicates, the overall contribution of medical care to health is rather modest at the margin, what determines marginal improvements in health? The answer lies in the other factors associated with the production of health, with education, income, lifestyle, and the environment being the major contributing factors.7

Education and Health  The positive relation between education and health is well documented in the literature. For example, Elo and Preston (1996) find that education had a significant impact on mortality for both men and women in the United States during the early 1980s, with the impact of education greater for men and those of working age than for women and the elderly. Lleras-Muney (2001) finds a significant relation between education levels and health. In particular, she finds that one more year of schooling decreases the probability of dying within 10 years by 3.6 percent. More recently, Cutler and Lleras-Muney (2006) estimate that an additional year of education increases life expectancy by between 0.18 and 0.6 years.

6. Except for the RAND study, which represents a social experiment, the other studies mentioned above are observational studies. Appendix A-1 points out that an observational study may only show an association rather than causation between two variables. Freeman et al. (2008) survey the literature and find fourteen studies analyzing the “causal” effect of health insurance on the utilization of medical services and health outcomes among nonelderly adults. Causality is likely because these studies use fixed effects, instrumental variables, or quasi-experimental approaches. Their review consistently shows that health insurance increases physician and preventive services, improves self-reported health status, and lowers mortality conditioned on injury and disease. Thus, these studies clearly show that the marginal product of medical care is positive. Unfortunately, the studies offer no direct estimates of the magnitudes of the marginal productivity of medical care among nonelderly adults. 7. A discussion of the impact of technology on health is postponed until Chapter 3.
Empirical studies have also documented a positive connection between income and health. Ettner (1996) finds that increases in income enhance both mental and physical health, while Lantz et al. (2001) find that income and education are both associated with improved health. More specifically, they find that people with less than a high school education and incomes below $10,000 are between two and three times more likely to have functional limitations and poorer self-rated health than their more advantaged counterparts.

While the positive relation between income and health is well established in the literature, a question remains concerning how temporary changes in the macroeconomy impact health. In other terms, what is the relationship between cyclical changes in the macroeconomy and overall health? Your first inclination is to assume that a procyclical relationship holds between the state of the economy and health. In other words, as an economy emerges from a recession and the unemployment rate begins to fall, overall health should improve. You might argue that higher per capita incomes should translate into improved health as people have more discretionary income to spend on medical care. In addition, as more people acquire jobs with employer-financed health insurance, the out-of-pocket price of medical care should drop, causing people to consume more health care. An improved economy may also be associated with healthier lifestyles because as unemployed workers find employment, stress levels are likely to fall along with alcohol consumption and smoking.

Ruhm (2000, 2003) argues that just the opposite may occur: an improved economy may be linked to poorer health. He cites three reasons why health may decline during a cyclical economic expansion. First, the opportunity cost of time is likely to increase with an improved economy. As workers find employment, the amount of leisure time they have to perform what Ruhm refers to as health-producing activities (such as exercise and eating right) diminishes. Second, the act of work may adversely impact the production of health. As the economy improves and more workers find employment, the number of work-related accidents and work-related stress cases increases. Third, an economic expansion may cause an increase in other causes of mortality such as traffic fatalities, homicide, and suicides.

To test the relationship between cyclical conditions and health, Ruhm estimates the impact that various economic indicators such as unemployment and personal income have on a number of health indicators. The author utilizes a state-based data set for the years 1972 through 1991 and estimates a number of equations utilizing a variety of health measures. Among the measures of health included in the analysis were overall mortality rates, age-based mortality rates, and deaths due to specific causes such as cardiovascular diseases, chronic liver disease and cirrhosis of the liver, motor vehicle accidents, and suicide.

The results are illuminating and suggest an inverse relationship between the strength of the economy and health in the short run. Overall, Ruhm finds that a 1 percent drop in the unemployment rate, relative to the state historical average, results in an increase in the total mortality rate of between 0.5 and 0.6 percent. In addition, Ruhm finds that the impact of changes in the unemployment rate on mortality rates appears to concentrate among the relatively young, between ages 20 and 44. This makes intuitive sense given they are the ones likely to be hit hardest by temporary changes in economic conditions.
The author also finds fluctuations in state unemployment rates to be inversely related to a number of specific causes of death. For example, Ruhm finds decreases in state unemployment rates to be associated with increased fatalities from auto accidents, other types of accidents, homicides, cardiovascular disease, and influenza. Ruhm (2003) also finds that a one-percentage-point decrease in the unemployment rate is associated with acute morbidity and ischemic heart disease increases of 1.5 and 4.3 percent, respectively. Ruhm’s empirical results are compelling because they suggest that cyclical, or temporary, changes in economic activity inversely impact health.

**Income Inequality and Health**  Lynch et al. (2004) and Wagstaff and van Doorslaer (2000) provide two excellent reviews of the literature regarding the relation between income inequality and health. Both papers agree that there is significant support in the literature for the absolute income hypothesis. The same cannot be said for the other alternative hypotheses, however. According to Wagstaff and van Doorslaer, there is “no support for the relative-income hypotheses and little or no support for the income-inequality hypothesis” (p. 543). They conclude that there is no empirical support for the relative position hypothesis. These results were largely reaffirmed by Lynch et al. (2004) and Lorgelly and Lindley (2008). However, Lynch et al., (2004) find some support for the hypothesis that greater income inequality worsens health outcomes at the state level in the United States.

**Lifestyle and Health**  The literature abounds with studies that illustrate the important role lifestyle plays in determining health. Among the risky lifestyle behaviors found to negatively impact health are smoking, excessive alcohol consumption, lack of physical activity, and poor diet. For example, Leigh and Fries (1992) estimate that the typical one-pack-a-day smoker experiences 10.9 more sick days every six months than comparable nonsmokers, while a person who consumes two or more drinks a day has 4.6 more sick days than a comparable light drinker (one or fewer drinks a day). Strum (2002) analyzes the impact of obesity, being overweight, smoking, and problem drinking on health and the consumption of health care for a sample of adults between ages 18 and 65 in 1997–1998. He finds that all four risk behaviors impact health to some degree, with obesity having the greatest impact. In fact, Strum estimates that obesity has the same impact on health as 20 years of aging when health status is measured by the number of seventeen common chronic conditions present. Finally, Balia and Jones (2008) find that lifestyle, particularly smoking and sleep patterns, play a significant role in predicting mortality. Using some rather sophisticated modeling and econometric techniques that focus on the distribution of health inequality, they estimate that predicted mortality rates may be much more sensitive to lifestyle factors, and less sensitive to socioeconomic factors and aging, than previously thought.

In the context of Figure 2–6, these results collectively suggest that adverse lifestyles cause the total product curve for medical care to shift downward and possibly flatten out. To compensate for the loss in health, a person may opt to slide up the total product curve by consuming more medical care. For example, Strum (2002) finds that obesity is related to an average increase in expenditures on inpatient and ambulatory care of $395 per year.

**Environment and Health**  The relation between environmental factors and health is mixed and, as a result, it is difficult to draw overall conclusions from the literature. Auster et al.
(1969) included two variables in the regression equation to capture the impact of environmental factors on health: an index of industrialization and a variable that measured the extent of urbanization. Both measures were hypothesized to be positively associated with such factors as air and water pollution, and therefore negatively related to health. The index of industrialization was found to cause higher mortality, but the level of urbanization had no influence.

Hadley (1982) undertook one of the more comprehensive assessments of the impact of environmental factors on health. Included in the regression analysis were variables representing water quality, air quality, climate, and occupational hazards. The results are inconclusive, which Hadley attributes mainly to “the lack of good quality data” (p. 73).

Other Determinants and Health Other variables found to contribute to health are age and marital status. The impact of marital status on health is interesting and merits a brief discussion. Married adults appear to experience better health than their single counterparts, everything else held constant. Most likely, this is because a spouse augments the production of health within the home. Marriage may also have a positive effect on health by altering preferences for risky behavior. Manor et al. (2000) find the mortality rate of married women to be lower than unmarried women for a sample of Israeli adult women, while more recently Kravdal (2001) finds that married people have a higher chance of survival of twelve common forms of cancer in Norway than their unmarried counterparts.

The Determinants of Health among Children

Numerous studies have investigated the factors that influence health among children. This body of literature is important because it illustrates the lasting impact of childhood health into adulthood. For example, Case et al. (2005) find that childhood health has a long-term impact on adult health, education, and social status. Such information is valuable when crafting public policies aimed at improving overall health.

Employing county-level data, Corman and Grossman (1985) regress the neonatal mortality rates for blacks and whites on a host of factors including education of the mother, the prevalence of poverty (a measure of income), and the availability of public programs. Some of the public programs included in the analysis are the existence of neonatal intensive care facilities, the availability of abortion services, organized family planning, and Medicaid. Overall, the results are robust and enlightening. Lack of schooling and the existence of poverty are found to raise the neonatal mortality rate for both white and black infants. Together, they account for an increase in neonatal mortality rates by 0.950 and 0.786 per 1,000 live births for whites and blacks, respectively. Access to health care also plays a role, as the presence of neonatal intensive care has caused the neonatal mortality rate to fall by 0.631 and 0.426 per 1,000 live births for white and black infants, respectively. Moreover, the results indicate that various government programs are associated with a reduced mortality rate for black as well as white infants. For example, Medicaid accounts for a decrease in the mortality rate by 0.632 per 1,000 live births for white children and 0.359 per 1,000 live births for black children.

8. The infant mortality rate equals the number of deaths from the first to the 364th day of life per 1,000 live births. The neonatal mortality rate represents the number of deaths from the first to the 27th day of life per 1,000 live births.
Two recent articles point to the significance of environmental factors on infant health. Chay and Greenstone (2003) use county data from 1981–1982 to estimate the impact of total suspended particulates (TSPs) on infant mortality. TSPs are minute pieces of dust, soot, dirt, ash, smoke, liquid vapor, or other matter in the atmosphere that can cause lung and heart disease. The authors find that a 1 percent reduction in TPS causes the infant mortality rate to fall by 0.35 percent at the county level. Currie and Neidell (2005) find that reductions in carbon monoxide also impact infant mortality. In particular, they find that reductions in carbon monoxide in California throughout the 1990s saved approximately 1000 infant lives. These studies are part of a growing body of literature that illustrates the importance of environmental factors in determining health.

Case et al. (2002) focus on the impact of socioeconomic status on children’s health.9 To no one’s surprise, the authors find a strong positive relation between the education of the parents and the health of their children. For example, the health of children is positively related to the education of mothers for children living with a mother. Education, in this case, is measured by whether the mother did not complete high school, had a high diploma, or had more than a high school education. The education of fathers is also found to positively contribute to improved health among children, implying that parental education positively impacts the production of a child’s health at all age levels.

The study also finds that household income is a strong predictor of children’s health. More specifically, the authors find that when household income doubles, the probability that a child 3 years old or younger is in excellent or very good health increases by 4 percent. Comparable improvements for children between ages 4 and 8, 9 and 12, and 13 and 17 are 4.9 percent, 5.9 percent, and 7.2 percent, respectively. Just as interesting, the authors find that permanent income is a strong determinant of children’s health. In particular, they find that family income before a child is born is positively related to the child’s health for all ages.

Finally, the authors find that healthier parents tend to have healthier children. Why that is the case, however, remains to be determined. However, the authors do estimate a series of equations for children with adoptive and biological parents and find that the impact of income on health is not significantly different across the two populations. While this evidence is not definitive, it does suggest that genetics may explain only part of the reason why healthier parents have healthier children. Could it be that the production of health takes place at the household level and that healthier parents are simply more efficient producers of health for all members of the household? Clearly, more research needs to be done before we fully understand how parental behavior coupled with socioeconomic factors impacts children’s health.

The literature concerning uninsured versus insured status and health outcomes offers additional insights into the effect of medical care on infant health as well as on other groups. However, we couch the discussion in terms of the relation between medical care and health because the only plausible pathway from insurance to health outcomes is through medical care (Levy and Meltzer, 2001). In a series of articles, Currie and Gruber (1996a, 1996b, and 1997), using a quasi-experimental design, examine the expansion of Medicaid eligibility by Congress on birth-related health outcomes. The authors exploit the

fact some states expanded Medicaid eligibility more than others did and at different times. By correlating the magnitude and timing of eligibility expansions with the magnitude and timing of changes in health outcomes, it is possible to determine if a causal effect of insurance on health holds. Currie and Gruber conclude that a significant increase in health inputs and a corresponding reduction in low infant birth weight and child mortality relative to a baseline results from an expansion in Medicaid eligibility. They also find that the magnitude of the Medicaid expansion’s impact on infant mortality depends upon the proximity of high-tech hospitals.

As another example, Hanratty (1996) examines the impact of Canada’s national health insurance program on infant health outcomes. Her identification strategy involves the fact that Canadian provinces adopted national health insurance at different times between 1962 and 1972. She observed changes in the mortality and birth weights of infants across Canadian counties at different introduction dates for the national health insurance program while controlling for other nonmedical determinants of infant health. Her results suggest a significant reduction in the infant mortality rate and a smaller reduction in the low birth weight rate after the introduction of national health insurance in the various provinces of Canada.

The Determinants of Health among the Elderly

Several studies have examined the medical care utilization and health of individuals who suddenly become Medicare-eligible at age 65 but previously uninsured to otherwise comparable individuals who were continuously insured. Lichtenberg (2002) analyzes the effect of Medicare on the health of elderly individuals by looking for sudden discontinuities in medical care utilization and health outcomes at age 65, when people typically become eligible for the federal program. Notice that chronological age is an external factor that cannot be altered by the nonmedical determinants of health or influenced by health status. He finds evidence that the utilization of ambulatory and inpatient care increases sharply at age 65. Lichtenberg also finds evidence that people spend less time in bed and face a reduced probability of dying compared to what would have occurred in the absence of Medicare. His results suggest a relatively large marginal productivity of medical care on the health of elderly individuals.

These results are reaffirmed by Card et al. (2007). Using data between 1992 and 2002, they examine the mortality rates of 400,000 elderly patients who were discharged from California hospitals before and after their 65th birthday when they become eligible for Medicare. To control for the possibility that some of the elderly may postpone medical care until they become eligible for Medicare, the authors compare Medicare-eligible people to uninsured individuals who were admitted to the emergency room for medical conditions that require immediate attention. Card et al. find that Medicare eligibility is associated with more medical spending and procedures and a reduction in the mortality rate of elderly individuals.

Using a nationally representative data set, McWilliams et al. (2007) provide a quasi-experimental analysis of longitudinal data for 5,006 adults who were continuously insured and 2,227 adults who were persistently or intermittently uninsured. Individuals ranged from 55 to 64 years of age. The authors find that acquisition of Medicare coverage is associated
with improved trends in self-reported health for previously uninsured adults, particularly for those with cardiovascular disease or diabetes.\textsuperscript{10}

The Role of Public Health: An Historical Approach

Thus far our discussion has revolved around the production of good health at the micro, or individual, level. Recall that the health production function, as specified in equation 2–2, is taken from the perspective of the individual in terms of the various inputs needed to produce health. We cannot ignore, however, the tremendous impact improvements in public health have had on health over time through an impact on the environmental and technology factors in equation 2–2. Public health places the emphasis on improving health at the community level and looks to such things as improving health education, controlling communicable diseases, improving sanitation, and monitoring and controlling environmental hazards. The fact that almost every municipality, county, and state in the country has a department of public health attests to the importance of public health on our everyday lives.

To illustrate the importance of public health, we discuss two very important public health interventions in the United States. The first health intervention deals with the development of clear water in the United States during the first half of the twentieth century. It coincides in our history with a number of improvements in nutrition and public health that caused infectious-disease mortality rates to decrease significantly. The second intervention deals with the development of a polio vaccine, which corresponds with the growth in modern medicine in United States starting in the 1930s with the development of sulfa drugs, or antibiotics. (Cutler, 2006)

During the first part of the twentieth century the United States witnessed an almost unprecedented advancement in health as measured by a drop in the overall mortality rate. Cutler and Miller (2005) provide a compelling case that a majority of this decrease in the mortality rate can be attributed to improvements in water quality brought about by public investments in clean water technologies. Their study uses historical data for thirteen cities where dates were available for four clean water interventions: water filtration, water chlorination, sewage treatment, and sewage chlorination. The dependent variables in the study include alternative measures of mortality. The empirical results suggest that improvements in water quality could explain 43 percent of the reduction in mortality rates from 1900 through 1936 across the cities in the sample. Even more convincing, cleaner water explained 62 percent of the drop in infant mortality and 74 percent of the decline in child mortality over the same time period.

Poliomyelitis, or polio, was one of the most dreaded epidemics to hit the United States in the mid-twentieth century. It is a highly infectious virus that generally afflicts children and can lead to paralysis or death. The most celebrated case occurred in 1921 when Franklin Delano Roosevelt, then a relatively unknown politician from New York, contracted polio

\textsuperscript{10} However, Finkelstein and McKnight (2005) find that the introduction of Medicare in 1965 had no measurable impact on elderly mortality during the first decade of the program. That is probably because many high-powered medical technologies such as angioplasty and stents were not available at that time. Finkelstein and McKnight did find the Medicare substantially reduced the exposure of the elderly to the out-of-pocket costs of medical care. Thus, while not initially reducing mortality, Medicare did offer a substantial amount of utility for elderly individuals because of the greater financial security.
while vacationing with his family. The disease left his legs paralyzed and he was largely wheelchair bound for the remainder of his life. While his disability was not hidden from the public, reporters were discouraged from taking pictures of him in his wheelchair while he was the governor of New York and later the president of the United States.

While polio had been around for many years, the number of new polio cases began to accelerate in the United States in the 1940s and early 1950s, reaching epidemic proportions in 1952 with 21,000 new cases. In 1955 the American public received news that Jonas Salk had developed a polio vaccine. The news was received nationally with much fanfare and Salk became a national hero overnight. With the support from the federal government and the March of Dimes, a plan was developed to distribute the vaccination across the country with priority given to young children. Within two years the number of reported polio cases fell by approximately 90 percent (Oshinsky, 2005).

This public health intervention is rather extraordinary because for the first time in our history a private philanthropic organization played a vital role in eradicating a major health problem. Much of the medical research and distribution of the vaccine was funded by the National Foundation for Infantile Paralysis, or the March of Dimes, which was started in 1938. Support for the foundation in terms of volunteers and funds was unprecedented and in 1954 alone the foundation raised an excess of $66 million.¹¹

The polio vaccine has improved over the years. Today states require students in licensed day care or kindergarten to be immunized for polio, with few exceptions. In many communities, local public health departments, or school clinics, provide vaccinations free of charge for those families who cannot afford to be vaccinated by a private health care provider.

These two examples illustrate the significant impact public health has had on reducing infectious diseases in the United States in the twentieth century. In the context of the total product curve, both public health initiatives caused the curve to shift and rotate upward as illustrated in Figure 2-5. Enhanced water sanitation improved the physical environment, while the polio vaccination is an example of a new medical technology. Needless to say, public health can impact the production of health in a variety of ways. Other examples may include a state-wide anti-smoking campaign aimed at improving lifestyle or a teenage pregnancy prevention program in the local high schools directed at enhancing sex education.

The Ten Major Causes of Death in the United States in 2005

As mentioned previously, individual choices, socioeconomic status, and environmental factors play a significant role in the production of health. If so, one might suspect that national disease-specific mortality rates would reflect the importance of these variables. That is, mortality rates should be high for diseases that are more sensitive to adverse lifestyles, low socioeconomic status, or unhealthy environments. With this in mind, Table 2-1 lists the top ten causes of death in the United States for 2005. Over the course of the year, more than 2.4 million individuals died in the United States. Of this number, approximately 77 percent succumbed to the ten most common causes of death listed in the table. By far

¹¹. Some of our readers may remember as a young school child being asked to donate a shiny new Roosevelt dime to the March of Dimes to help eradicate polio.
the number one cause of death is diseases of the heart, accounting for almost 27 percent of all deaths in the United States in 2005. Although researchers are still unclear as to what determines an individual’s risk for heart disease, they are certain that the blood level of cholesterol, smoking, level of physical activity, stress, and obesity play a major role in determining the risk of heart disease. Each of these factors is influenced by lifestyle choices, socioeconomic status, and environmental settings.

The second leading cause of death is malignant neoplasms, or cancers. Lifestyle choices often have an impact on this type of illness as well. For example, Edlin and Golanty (1988) point out that approximately 80 percent of all lung cancer deaths, the most common form of cancer, can be attributed to smoking. Socioeconomic status and environmental factors also come into play in determining the likelihood of contracting lung cancer through exposure to such items as asbestos and radon. The third leading cause of death is stroke and the medical community is in agreement that lifestyle, such as whether a person follows a proper diet and exercises, impacts the chances of having a stroke.

The fourth leading cause of death is chronic lower respiratory diseases, which includes chronic obstructive pulmonary disease, emphysema, and chronic bronchitis. Air pollution plays a critical role in the progression of these diseases. The next leading cause of death is unintentional injuries, which deals with deaths directly related to individual behavior such as automobile and industrial accidents rather than natural causes.

Finally, the list is interesting for what it does not include. In 1995 the human immunodeficiency virus (HIV) was the eighth leading cause of death and accounted for 32,655 deaths. By 2005 that number had dropped to 12,543. This dramatic decrease in the number

---

**TABLE 2–1**
The Ten Leading Causes of Death in the United States in 2005

<table>
<thead>
<tr>
<th>Cause</th>
<th>Number of Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Diseases of the heart</td>
<td>652,091</td>
</tr>
<tr>
<td>2. Malignant neoplasms</td>
<td>559,312</td>
</tr>
<tr>
<td>3. Cerebrovascular diseases (stroke)</td>
<td>143,579</td>
</tr>
<tr>
<td>4. Chronic lower respiratory diseases</td>
<td>130,933</td>
</tr>
<tr>
<td>5. Unintentional injuries</td>
<td>117,809</td>
</tr>
<tr>
<td>6. Diabetes mellitus</td>
<td>75,119</td>
</tr>
<tr>
<td>7. Alzheimer’s disease</td>
<td>71,599</td>
</tr>
<tr>
<td>8. Influenza and pneumonia</td>
<td>63,001</td>
</tr>
<tr>
<td>9. Nephritis, nephritic syndrome, and nephrosis (kidney disease)</td>
<td>43,901</td>
</tr>
<tr>
<td>10. Septicemia</td>
<td>34,136</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2,448,017</td>
</tr>
</tbody>
</table>

of deaths can be attributed to a series of factors including improved therapies and changes in lifestyle brought about by great public awareness of the disease.

This rather simple exercise underscores the importance that lifestyle choices, socioeconomic status, and environmental factors play in determining deaths in the United States. It is worth noting that the information in Table 2–1 can also be used to illustrate the importance that an individual’s mental and physical profile play in the making of health. For example, age is a critical factor in determining the onset of Alzheimer’s disease, while the environment, genetics, and age contribute to development of diabetes.

**Empirical Evidence on the Production of Health: A Summary**

Health production theory suggests that medical care, lifestyle factors, environmental surroundings, and socioeconomic status all influence health conditioned upon the state of medical technology and an individual’s medical profile. Clearly, the total impact of medical care on health is significant and many people would die without proper medical care attention. But from a practical economic perspective, it is important to know which factors contribute more to improved health at the margin so cost-effective policies can be designed. Given limited resources, society’s goal is to implement least-cost methods of improving population health.

In terms of adult health, evidence seems to suggest that nonmedical factors generate greater improvements in health at the margin than medical care. A better lifestyle and improved socioeconomic and environmental conditions seem to matter more than the consumption of additional medical care. Medical care appears to be more important at the margin for infants than adults, especially for low-income infants. But as we saw in this chapter, socioeconomic and environmental conditions are also important for infant health. In fact, even lifestyle is important for infants. While at first blush that statement may sound odd, low birth weight and greater infant mortality have been linked to adverse maternal lifestyle behaviors such as tobacco, alcohol and drug abuse. For the elderly, particularly those without health insurance prior to becoming Medicare-eligible, medical care is also important at the margin. But even in this case, nonmedical factors, such as exercise and diet, play an important role.

These empirical findings have some rather interesting policy implications. They suggest that any public policy initiative aimed at improving health should first consider raising education levels, reducing the amount of poverty, and encouraging improved lifestyles rather than simply providing additional medical care. Naturally, the specifics of any policy should be based on sound cost-benefit analysis.

**Summary**

Health, like any other good or service, is desired because it generates utility. Also like other goods and services, health is subject to the law of diminishing marginal utility. This law stipulates that each additional unit of health provides less marginal utility than the previous unit.

The making, or production, of health is influenced by a variety of factors, including the amount of medical care consumed. The positive relation between health and medical care, however, is nonlinear due to the law of diminishing marginal productivity. This law
underlies a fundamental production relation stating that health increases at a decreasing rate with additional amounts of medical care, holding other inputs constant. Some of the other factors determining health are the state of medical technology, the individual’s initial health profile, socioeconomic status, lifestyle, and environmental factors.

The empirical evidence for adults indicates that good health depends only moderately on the consumption of medical care. Socioeconomic status and lifestyle appear to play a much greater role in the production of good health of adults. Health appears to be more sensitive to changes in the consumption of medical care for vulnerable segments of the population such as low-income young and elderly individuals. Historically, health improved in the United States in large part because the number of deaths from infectious diseases decreased because of advances in public health.

Review Questions and Problems

1. Describe the factors that make it difficult to measure output in medical care markets.
2. As mentioned at the beginning of the chapter, the life expectancy rate in Russia fell significantly from 1989 through 1994. Use health production theory to explain what would happen to the relationship between good health and medical care in Russia if alcohol consumption diminished and the market economy strengthened. Provide a graph to illustrate your explanation.
3. Use health production theory to explain the role gender plays in the production of health during pregnancy. Provide a graph to illustrate your answer.
4. Use production theory to graphically illustrate the case in which a medical innovation improves health without any change in the consumption of medical care.
5. In your own words, use utility analysis to explain why people demand health. How does the law of diminishing marginal utility fit into the analysis?
6. Explain how an increase in income would affect the level of health in a relatively affluent country like the United States compared to a relatively poor country like Haiti.
7. You have just been appointed to the post of surgeon general of the United States. The president wants you to develop an advertising campaign called “A Healthy America by the Year 2020” that encourages Americans to lead a healthier lifestyle. What types of behavior would you try to influence? Why?
8. You have just been hired by a major metropolitan city as a health policy analyst. Your assignment is to devise a plan that city authorities could implement to lower the infant mortality rate. Based on the results cited in this chapter, what types of policies would you recommend? Substantiate your answer.
9. Explain how a change in each of the following factors would alter the shape of the total product curve for medical care.
   A. An increase in education.
   B. An improvement in lifestyle.
   C. An improvement in the environment.
10. Some people believe that cigarette and alcohol advertisements should be banned completely in the United States. If this were the case, what would likely happen to the shapes of the total and marginal product curves for medical care?
11. Explain why a researcher must be careful when interpreting findings from a survey that finds a positive association between education levels and health outcomes.

12. Consult the website of your state or county Public Health Department. Are there any public policy initiatives currently in place aimed at improving lifestyles, enhancing access to health care, or impacting the environment? Explain the intent of these policies in the context of production theory.

13. In a 1991 issue of the *Cato Journal*, Santerre, Grubaugh, and Stollar estimate an infant mortality equation using a sample of 20 countries belonging to the Organization for Economic Cooperation and Development (OECD) during the 6 adjacent half decades from 1960 to 1985 and a fixed effects model. They obtained the following (abbreviated) results:

\[
\text{IMR} = 3.93 - 0.069\text{TIME} - 0.892\text{RGDP} - 0.539\text{PHYS} + 0.707^{*}\text{URBAN} - 0.004\text{FLFPR} - 0.135\text{ED}
\]

\[
(2.60) (1.11) (6.83) (6.89) (4.21) (1.21) (2.34)
\]

Adjusted R^2 = .954 N = 110

All of the variables have been converted to logarithms so the coefficient estimates can be treated as elasticities. The numbers below the estimated coefficients represent t-statistics.

- IMR = infant mortality rate in each country for each year
- TIME = a time trend from 1 to 5 (1960 to 1985) capturing changing technology and knowledge
- RGDP = real gross domestic product per capita in each country for each year
- PHYS = number of physicians per capita in each country for each year
- URBAN = percentage of the population in urban areas in each country for each year
- FLFPR = female labor force participation rate in each country during for year
- ED = level of education in each country for each year.

Based upon these findings answer the following questions:

A. What percentage of the variation in the infant mortality rate is explained by the independent variables? How do you know that?

B. Using health production theory as much as possible, provide a hypothesis or theory about the relationship (direct or inverse) between the first three independent variables and the infant mortality rate.

C. Are those three hypotheses supported by the regression results? Explain.

D. Given that the estimated coefficients are also elasticities, interpret the coefficients on the number of physicians and real GDP.

E. Should we expect the physician elasticity to remain constant if increasingly more physicians are employed in the typical health economy? Why or why not?

F. Based upon those findings explain why the infant mortality rate may be so much higher in Turkey than Japan?

**Online Resources**

To access Internet links related to the topics in this chapter, please visit our web site at [www.cengage.com/economics/santerre](http://www.cengage.com/economics/santerre).
References


CHAPTER 2 Health and Medical Care: An Economic Perspective


