Getting Health Reform Right

By

Marc J. Roberts
William Hsiao
Peter Berman
Michael R. Reich

October 2001

Table of Contents

Preface

This book represents the culmination of four years of intense collaboration and cooperation among the authors. In the deepest sense, it is a joint product: the result of innumerable meetings (lasting from two hours to two days), notes and memos containing suggestions and ideas, conversations and confrontations – ranging from the delightful to the heated. Chapters have been drafted and redrafted, edited and critiqued and drafted again. We each brought different skills and capacities to our work, some did more writing, others more critiquing. Some were especially good at conceptualizing, others had insights and wisdom derived from years of hard-won experience. No two or three of us could have produced what the four of us have produced together.

In this preface we explain the process of producing the book and acknowledge the inspiration and help of many people.

How did four individuals with disparate training and experience come to cooperate on a book on health sector reform? The credit for initiating the project has to go to Paul Shaw at the World Bank Institute. In 1996, Shaw organized a major teaching program at the Bank on health sector reform and financing. He organized an advisory group that included Berman and Hsiao and from his efforts emerged the design of what became known as the "Flagship Course on Health Sector Reform and Sustainable Financing." The course included a series of modules, and Shaw asked Hsiao to take responsibility for organizing an introductory module on health systems assessment and diagnosis. Hsiao, acutely aware of the multidisciplinary nature of the task, recruited Berman, Reich and Roberts to join the team, and our collaboration began.

In the summer of 1997, we produced a six-chapter background note for the Flagship Course, along with various teaching materials. The course included modules developed by various groups around the world, and took place in Washington, DC, with 90 participants from many countries. Our background papers, entitled "Diagnostic Approaches in Assessing Health Care Systems," became the basis for this book. After the course was over, we reviewed our materials and agreed that much more work was required. In the summer of 1998, the Bank retained Roberts to rewrite those materials – and the resulting draft moved us another step forward.

Realizing that a great deal more work was needed, the four of us agreed to collaborate in turning the background papers into a book manuscript. The subsequent multi-year effort was financed in part by a grant to Reich from the Edna McConnell Clark Foundation. In addition, we received continuing support from the Bank to each of us for preparing and teaching in the Flagship course in Washington and in courses offered at partner institutions around the world. As we went along, all the chapters were written and re-written many times and the current manuscript emerged.

In Chapter 1 we identify six key conceptual contributions of this book; here we note briefly the initial source for each one.

The policy cycle formulation emerged from work that Roberts did with a colleague, Christian Koeck, for a course that they taught on health policy.

The ethical framework was developed by Roberts and Reich, together with a colleague, Karl Lauterbach, for a course they have taught on public health ethics for the past decade.

The political analysis approach used in this book was developed by Reich, along with accompanying software he has produced, called *Policymaker* (with David Cooper).

The concept that health systems are means to ends and *the performance criteria formulation* were developed by Hsiao from his research and advisory role to many countries. Roberts and Berman elaborated the core and intermediate criteria and the relationship to the ethical framework.

The control knob conceptualization was developed by Hsiao—with three of the specific knobs, organization, regulation and behavior, extensively deepened and expanded from our conversations.

The diagnostic tree approach came from Roberts.

All of these ideas play a role in the book that follows. We do want to stress, however, the critical conceptual armature around which our book is built. We consider this to be the core idea of the technical means-end perspective embodied in the control knobs and the performance criteria, with the broader perspective of ethics and politics as the inescapable (and legitimate) context within which problems are identified and solutions developed.

Now for some words about responsibility for the current draft. Chapters 1 (introduction), 2 (policy cycle), and 4 (core criteria) were written by Roberts with substantial substantive and editorial input from Reich. Chapter 3 (ethics) is by Roberts and Reich. Chapter 5 (politics) is by Reich. Chapter 6 (diagnosis and policy development) is by Berman and Roberts. Chapter 7 (financing) is by Hsiao and Roberts. Chapter 8 (payment) is by Hsiao, Chapter 9 (organization) by Roberts and Berman, Chapter 10 (regulation) is by Hsiao and textual input from Roberts. Chapter 11 (behavior) is by Reich. Chapter 12 (conclusion) is by all of us. Of course, each of us contributed important ideas to all of the chapters, not adequately reflected in this list.

A project of this complexity and duration is only possible with the support of many people. Our greatest debt is to Anne Johansen of the World Bank. As the organizer of many of the courses in which we have taught, and as a teacher herself in the courses, she has contributed many crucial insights over the years. She has pushed us repeatedly to clarify our ideas—and not infrequently made useful suggestions as to how we might do so. Paul Shaw, who began this whole effort, and Hadia Karam, also of the World Bank Institute, have likewise been valued critics, commentators, and partners. Others who have

taught in the course have contributed greatly to our thinking include Ricardo Bitran, Alan Maynard, Alex Preker, George Schieber, and Melitta Jakob. We also wish to thank our colleague at Harvard, Tom Bossert, whose work on decentralization we drew on extensively, and who provided thoughtful comments on several chapters. Our editor, Donald Halstead, improved our writing and continuity with good will and graciousness. Several people helped over the years with typing and revisions, including Katrina Meyer, Betsy Barker, Vanessa Bingham, and Margaret Ou. We also appreciate the support from colleagues who taught with us – especially Miklos Soska and Tomas Etavitz in Budapest. Finally, we want to thank the literally hundreds of participants in Flagship courses both in Washington and around the world. Their energy, ideas, suggestions, and responses were invaluable, as they struggled with various versions of this material. We hope that our efforts will help them with the vital work they do every day, trying to improve the functioning of health care systems around the world.

Chapter 1
Setting the Scene

1. Introduction

Throughout the world, governments are engaged in health sector reform. The transitional economies of Eastern Europe are full of new social insurance schemes. Nations in South America are experimenting with ways to extend insurance coverage to both the rural and urban poor. In Africa, experiments with fiscal decentralization have produced additional revenues for hospitals, but also additional inequality between rich and poor regions. To improve efficiency, many nations are also experimenting with both new payment systems and new ways to organize health care delivery.

Too often, however, conflicting political calculations, economic implications, and ethical concerns have led to a confused national debate: How can we deal with doctors' demands for more money? What strategies exist to reduce costs for medical care while expanding social insurance to cover the poor? Should we expand the system of publicly provided health centers, or move more to private practice family physicians? Should we ask patients to pay more out of pocket, or make more use of general tax revenues? Is the answer more new technology or less? More doctors or fewer medical schools? Building new hospitals or spending more on anti-smoking campaigns?

This book is intended to help health reformers develop the skills they need to answer these questions. It introduces a set of concepts that can facilitate systematic and critical thinking about health sector reform. Our experience in many countries is that such careful analysis is both possible and potentially very useful, as we hope to demonstrate in the pages that follow.

Our discussion draws on a range of disciplines. For an understanding of how money is raised and spent, we make extensive use of economic analysis. For insight into how individuals and organizations react to those incentives, we call on sociology, psychology, and organizational theory. For guidance on how to get programs adopted in the real world of government decision-making, we rely on political analysis. In addition, for thinking about how to define problems, we devote significant attention to moral philosophy. Our criteria for the use of theory have been pragmatic: Which conceptual tools and ways of thinking can help real people succeed at the real work of health sector reform?

Simply reading a book about how to do health sector reform cannot prepare someone for the work at hand. On the contrary, much of what is required resembles a skill or a craft, like cooking, kicking a football, sailing a small boat, or playing the trumpet. Such skills are best developed through supervised practice. The materials presented here, therefore, need to be supplemented by active discussion, both of case examples and personal experiences. Applying these concepts to specific cases reveals how the ideas are both helpful for solving particular problems and yet also have their

limits. Such explorations can develop the intuition and judgment that are needed for successful action in a particular context.

The goal of this book is to develop a more reasoned and effective approach to improving the performance of health care systems. The ideas and methods are based on our own engagement with health reform in many parts of the world, in countries with different types of problems and different attempts at solutions, not all of which have been successful. We have also taught this approach to of practitioners around the world, from Malaysia, Kazakhstan and China, to Hungary, Lebanon and Russia, to Mexico, Chile and Washington, DC. In developing these materials over the course of several years, we have challenged each other to clarify ideas, and to make general arguments relevant to actual practice. The result is a book addressed primarily to practitioners – people concerned with making health reform happen – although we hope that researchers will also find it stimulating and provocative.

We want to stress that our approach is based on looking at the health care system as a *means to an end*. We can only know whether the system is working well or badly, or identify promising reforms, by keeping this perspective clearly in mind. Thus our method focuses on the need to identify goals explicitly, diagnosis causes of poor performance in a systematic way and devise reforms that will produce real changes in performance. We will argue again and again that reform must be strategic, based on honest means-ends analyses of what is likely to happen in a particular national context. Reforms need to be judged not on reformers' intentions, but by the changes they actually produce.

Too often, advocates of particular health sector reform ideas do not offer arguments that meet this burden. Instead they uncritically urge adoption of their favorite idea: be it decentralization or family medicine, primary care or private insurance. Yet they rarely identify the performance problems they want to improve, or how the reform they advocate will lead to that improvement.

We know from experience that health sector reform a very difficult process. As we discuss below (Chapter 5) existing institutions and interest groups often have both the reasons and the resources to vigorously oppose change. As a result, it often takes some sort of political or economic shock to begin the health sector reform process: a budget crisis, a change in the government coalition, a public scandal, a strike by providers or some combination of these and other similar events. This means that major changes in any country's health sector are infrequent. Hence reformers have to be prepared to energetically seize the opportunity for major change when it occurs.

We are not naïve, however, in believing that a crisis will necessarily produce unity about what needs to be done. Differences in values, interests, political philosophy and institutional responsibility will all make themselves felt. The Ministry of Finance is not likely to agree with the Ministry of Health. The doctors will not necessarily see eye to eye with the hospital administrators; nor will local government leaders agree with pharmaceutical company executives. It is exactly for this reason that we stress the importance of being clear about underlying values – and the ways in which these values

lead to different goals and different reform priorities. For only by being clear about goals is it possible to devise a program that has any hope of achieving the changes that reformers desire.

Experience has also taught us that, because of its complexity, the behavior of a health care system is not easy to control. Change payment schemes, and doctors and hospitals will modify their behavior to defend their incomes. Impose regulations on hospitals, and reports will be adjusted to show compliance. Create new payroll taxes and some businesses are sure to try to avoid them. Moreover, the casual relationships in the system are complex. Change incentives to hospitals to foster efficiency, but not the authority of hospital managers, and the new scheme may produce little change. Institute competitive bidding in situations where there are few competitors, and the hoped for decline in prices and costs may not occur.

These characteristics of the system – its complexity, resistance to change and the diversity of perspective within it – tend to give health sector reform an episodic and cyclical character. When some internal or external shock does focus national attention on health sector reform, a specific feature (or features) of the system is often identified as critical. And this then becomes a target for major reform efforts. But the initial changes often lead to further unanticipated problems. And additional rounds of (often less dramatic) reform often occur. As a result, the initial changes are adapted, perfected, modified (or even disassembled) by subsequent actions. We discuss the health reform cycle in more detail in Chapter 2.

Moreover, exactly because we focus on consequences, we will insist throughout that matters of practicality and implementation must always be kept in mind. While international experience is a valuable source of ideas and guidance, we urge the reader to remember that policies that have worked well in other countries always need to be evaluated in your particular context. Every nation is different and simple imitation is seldom advisable. Instead reformers have many questions to ask themselves. How good are our data systems? How respected are our courts? How energetic are our administrators? This is why we said above that policy needs to be developed in a realistic and self-critical manner. Because only such skepticism and self-examination will lead to plans that have a reasonable prospect of success in a reformer's own national context.

Given this understanding of how reform occurs, our approach to the health sector reform process brings together six important elements:

- A description of the *policy cycle* and an identification of the key tasks a reformer should be prepared to tackle at each stage of the process (Chapter 2).

- A guide to *ethical theory* to help think through the moral basis for the varying policy goals and priorities that are the basis for defining the reform agenda (Chapter 3).

- A core set of *fundamental health system performance criteria and intermediate objectives*, which reformers can use as guides to measuring performance. We link these goals to ethical theory (Chapter 4).

- An introduction to systematic *political analysis*, because politics matters at each step of the health reform cycle. Political analysis should be done early and often throughout the policy cycle (Chapter 5).

- A *systematic approach to health system diagnosis* that can be taught and practiced effectively. It is based on the idea of working backwards from identified performance problems to their causes and the causes of causes, using a technique we call a diagnostic tree (Chapter 6).

- A framework of *five health system "control knobs"* that summarizes the options available to reformers (payment, financing, organization, regulation, and behavior) for influencing health system performance. These provide the focus for both the diagnostic process and for developing policies to achieve better performance (Chapters 7 to 11).

By combining these six sets of concepts, ideas and techniques, practitioners have a better chance of "getting health reform right."

However ambitious, this book does *not* cover everything about health sector reform. It does not provide an exhaustive discussion of a large number of detailed policy options. There simply is no room for us to cover so much material. Nor does it present a single set of preferred solutions, exactly because, as we explain shortly, countries are too varied for one approach to be universally applicable. Instead, we provide you, the reader, with new ways of talking and thinking about the problems you confront, so that your own analysis of your situation, and the decisions you make to deal with it, can be more effective.

In our view, the point that the problems and the solutions for health sector reform vary across nations is almost too obvious to need repeating. Yet many ignore this vital point. Countries differ widely in their levels of economic development, social conditions, disease patterns, and institutional arrangements. Nations range in per capita income from several hundred dollars per year to twenty thousand and more. In some countries half the children born alive die before the age of five, in others child death is an unusually tragic event. There are many countries where physicians are as rare as one per 10 or 20 thousand population, and not a few where there may be one per several hundred.

Other differences in national conditions are equally striking. In some countries, the civil service is reliable and energetic; in others it is corrupt and ineffective. Some countries are blessed with mineral or agricultural resources; others struggle with difficult climate or poor soils. Some suffer from endemic infectious diseases or the AIDS epidemic; others struggle with increasing chronic disease. Some countries have relatively stable political conditions with good governance; others have experienced years of

political terror and instability. These factors affect both how the problems of health sector reform are defined, and the kinds of solutions that are most likely to work.

Moreover, countries seek different goals with their health care systems. Countries have serious and legitimate differences over whether they seek to provide equal care to rich and poor, urban and rural, productive and unproductive. They differ over which health conditions they seek to address as high priorities: the burden of maternal and infant mortality, for example, versus conditions that affect those in their economically productive years. Countries also differ in their views about how forcefully citizens can be pushed to change their personal habits or their culture, when these support unhealthy behaviors. In short, not only do problems differ from country to country, but so do values, goals, and ethics. We believe, therefore, that developing a plan for health reform requires a direct engagement with philosophy and social values.

Finally, nations differ in their political systems. Some are elected democracies, others are one-party systems, while others are ruled by hereditary monarchs. Among the democracies, some constitutions centralize power, while others decentralize it. Some electoral systems encourage fringe parties, while others discourage them. Some political systems encourage the activities of non-governmental organizations, while others suppress any groups that might challenge the dominant power structure. In addition, different groups within a country vary in power and position, depending on their leadership and history. These differences in political structures, political practices, and political freedom affect the processes of health reform, including how problems and solutions are defined, and the kinds of policies that are politically feasible.

Thus, the analysis we offer of health reform is resolutely contextual – both problems and solutions depend on the particular situation. Our goal is to illuminate that dependence by presenting conceptual tools for analyzing problems and solutions within particular contexts. At the same time, we also recognize that common patterns in health reform exist: certain large forces operate, to varying degrees, in all countries. We briefly review these shortly. But before doing so, we need to make one other preliminary point.

No sharp line distinguishes what *is* and what *is not* part of the health care system, nor is it worth much time and energy debating about the system's boundary. We argue below that a pragmatic analysis must focus on the *consequences* of this system for the *well-being* of citizens. The analytical net must be cast wide enough to include all forces and factors that reformers might want to influence in seeking to improve the lives of their fellow citizens.

Bearing this caveat in mind we use the terms "the health sector" and "the health care system" in this book to include the following:

- All those who deliver health care – public or private, Western or traditional, licensed or unlicensed. This includes doctors, nurses, hospitals, clinics, pharmacies, village health workers, and traditional healers.

- The money flows that finance such care are also part of the system – official or unofficial, through intermediaries or directly out of patients' pockets.

- The activities of those who provide the specialized inputs into the health care process are also part of the system. This includes medical and nursing schools, as well as drug and device manufacturers.

- The financial intermediaries, planners, and regulators who control, fund, and influence those who provide care are also part of the system. This encompasses ministries of health, finance and planning, social and private insurance institutions, and regulatory bodies.

- The activities of organizations that deliver preventive services, ranging from family planning, to infectious disease control, to education for nutrition, smoking, and substance abuse. These may be public or private, local, national, or international.

We take a similar pragmatic view of what we mean by "reform." Reforms differ along at least two dimensions: (1) the number of different aspects of the health care system that are changed, and (2) how radically the individual changes depart from past practice. But as we will stress repeatedly (especially when discussing particular options in the book's second half) successful reform often involves introducing a set of inter-dependent and mutually supporting interventions – especially if a major departure from past practice is made in one piece of the system.

2. The Forces Driving Health Reform

Four broad forces are driving health reform in countries around the world. The first is *rising costs* in health care. In nearly every country, the costs of medical care are forcing governments to rethink their policies and approaches to the health system. In addition there are *rising expectations*, as citizens demand more, both from government in general and from the health care system in particular. These rising expectations are the second major driving force behind the attention to health reform in many countries.

Rising costs and higher expectations are occurring at a time in history when governments confront *limits on the capacity to pay* the costs of health care. These have arisen both for developing countries that have confronted political instability and economic turmoil, and countries that have experienced many years of relative peace and prosperity. These limits represent the third driving force at play.

The current worldwide reform debate is also influenced by a fourth driving force: growing *skepticism of conventional approaches* to the health sector. This skepticism takes many guises: calls for new organizational forms, dismay at poor governance and ineffective bureaucratic performance, investigations into corruption and inefficiency, and the questioning of old dogmas. For a reformer, this questioning provides both a problem

and an opportunity: a problem because more is being demanded and more options must be analyzed, and an opportunity because there is more openness to change and innovation.

The variations in national circumstances we noted above mean that the way in which these forces work themselves out varies a great deal from case to case. The world's poorest countries are especially vulnerable to external shocks, both economic and political. Their dependence on international financial institutions and markets means that significant health system changes may be requested or imposed as a condition of external assistance. In addition, the HIV/AIDS pandemic is placing immense strains on some of the world's nations with the least resources to meet that challenge.

Other lower-income nations are very concerned with the maldistribution of limited resources in the face of pressing health needs. In these nations, service levels and health status may vary greatly between the emerging urban middle class and the rural poor. In addition the public sector is often perceived as providing poor service, and many citizens – even among the poor—pay for services out of pocket. In still other middle- and lower-income nations health sector reform has been initiated as part of an orderly and intentional process of change. Citizens' demands for financial protection from high health care costs have led governments to seek financing strategies that will be sustainable as development proceeds.

Regardless of how the drama unfolds, the struggle by governments to resolve the contest between what they can spend and what they or their citizens want is often the central health sector reform dilemma. (We provide a more extensive analysis of how to think about such cost-performance dilemmas in Chapter 4.) With this variation in mind, however, it is still useful to review certain worldwide patterns that provide a broad context for many national reform efforts.

2.1. Force #1: Rising Costs

All around the world, health care costs are being pushed upward by changing demographics, evolving disease patterns and new technology, as well as by rising expectations. This pattern is best documented in the world's wealthier countries, where health spending has risen from an average of 5% of national income in 1970 to 8.1% in 1997, a 60% increase, with far from commensurate improvements in longevity (Huber, 1999). A richer, older, more secular and knowledgeable population wants more medical care, and there is ever more in the way of new drugs and devices to spend money on. The rise of chronic disease exacerbates the cost problem in several ways, influencing technology development and utilization, and troubling insurance markets.

A major force driving costs upward, especially in relatively more developed countries, is the aging of the population. For example, the percentage of the population aged 65 years and older is expected to increase from 6 percent to 18 percent in Asia between 2000 and 2050, and from 5.5 percent to 18 percent in Latin America over the same period (National Research Council, 2001). Birth rates have declined in many nations due to various causes, including increases in women's education and status,

increased availability of contraception and abortion, and higher child survival rates. In addition, urbanization and economic prosperity have changed views about desired family size.

However, even in some poorer countries where birth rates remain high, the proportion of the population that is young is still growing, and infectious disease is still a problem, a long history of rapid population growth means that the total number of older people is also often increasing. This can give rise to a dual burden of disease, as the nation has to cope with everything from childhood diarrheal disease to rising coronary artery disease rates.

To some extent, the increase in chronic disease is due to success of the health care system: more citizens now live long enough to develop chronic disease. If more people died sooner per capita health care costs would be lower. This is because as we age, our arteries harden, our joints grow stiff, our livers and kidneys work less efficiently, our hearts get tired, and our minds become cloudy. These changes create added demands on the health care system. Such "failures of success" are reflected in the *rising* levels of *disability* in most industrialized and many middle-income countries. For example, WHO estimates that, mainly due to population aging, the share of the total burden of disease accounted for by psychiatric and neurological conditions will increase from 10 to 15% between 1990 and 2020 (WHO, 1996). Ironically, countries with successful population control policies and health care systems move down this path further and faster. Indeed, the success of health care systems in lowering childhood mortality also increases the demand for services from that age group.

Life expectancy at birth has risen from about 40 years in 1950 to 63 years in 1990 and is still rising globally. Declining mortality rates are due to many factors. Economic growth leads to improved nutrition and housing, and lower physical stress and strain. Public health measures also contribute, especially clean water, better sewage disposal, immunization, and infectious disease control. But whatever the cause, the consequence of lowered mortality is an increase in the number of people who need medical care.

Other epidemiological developments are also driving health reform, as rising costs confront limited resources. The most notable example is AIDS, which threatens to overwhelm the health care system, and many other social institutions, in the most seriously affected countries. AIDS is now the leading cause of death for adult men in their middle years in severely affected countries and brings with it a major accompanying burden of ill health, disability, and family tragedy. In Eastern Europe, increases in alcoholism and suicide have produced serious systemic problems: in Russia and several republics of the former Soviet Union , male life expectancy has actually fallen ten to fifteen years in the last decade. Despite successes against smallpox and polio, other infectious diseases (from tuberculosis to dengue fever) are on the increase in various parts of the world. Such factors are of grave concern to health reformers.

The sellers of health care goods and services also contribute to rising costs by working hard to take advantage of market opportunities. It is a paradox that many poor

countries have large, private fee-for-service health sectors – a pattern of organization that tends to produce significantly higher expenditures. The pharmaceutical companies also respond to the market and now have compounds for every organ system – drugs that will regrow your hair, cure your depression, lower your cholesterol and blood pressure, combat your ulcers, and improve your sexual functioning. In fact, the middle class in industrialized nations is at risk of becoming a pharmacological layer cake, and efforts to market these compounds worldwide go forward every day. In poor countries, even traditional medicine is now becoming commodified, with modern packaging, marketing, and of course prices.

New technology is also driving up health care costs. The largest gains from innovation, economically and professionally, come from developing high-priced solutions to previously unsolved problems – Viagra sells for $10 per pill in the U.S. Nor is there much scientific prestige associated with developing small, cost-saving changes to existing technology. Solve a previously unsolved problem, however, and even if the therapy costs $50,000 per case, you might get nominated for the Nobel prize – or at least receive a professorship at a famous university. Technological progress can, for example, lower the unit cost of a routine test or provide a cheap drug for a condition now treated with expensive surgery. But even this may not always reduce total costs, if the new product unleashes a large unmet demand. For example, drugs for anti-retroviral therapy, which cost $15,000 or more per patient annually in higher income countries may now be available for $350 in lower income countries. But few expect that this drastic reduction in drug cost in Africa would reduce *total spending* on those drugs.

The increasing prevalence of chronic disease also disrupts the functioning of health insurance markets. Insurance exists to cover unpredictable risks. But in a world of chronic disease, many of next year's health care costs are predictable: those who will be sick in the future are often those who are already sick today. This creates perverse incentives on both sides of the health insurance market. Sellers of insurance try not to sell to sick people, or do so only at a very high price. Hence, the sick, who most need coverage, will often remain uncovered. Conversely, those who know they are healthy will have every incentive not to purchase insurance, because it is not a good buy for them, leaving only the sick in the insurance scheme and decreasing the amount of "risk pooling" in the system.

2.2. Force #2: Rising Expectations

The rising expectations of the health care system derive from three sources: economic, social, and political. When we compare spending across countries, health care is what economists term a "luxury" good – a good whose demand rises more quickly than income. Thus, paradoxically, recent worldwide prosperity is one cause of the problems that many governments confront within the health sector. As nations grow economically, their citizens want to spend a rising share of the nation's output on health. These rising expenditures represent the use of more, and more costly, services for both care and cure. In some regions of the world, like East Asia, which (until recently) had seen years of steady growth, the rise in expectations has been considerable.

Global social developments reinforce these demands. Films, television, and the Internet have given the citizens of many countries an image of life in the world's high-income nations. Such exposure has contributed to a rising materialism and loss of traditional values. As travel becomes less expensive, more individuals in the developing world are exposed to friends or family members who have returned from working or traveling in Europe or North America. Increasingly, people around the world accept that it is desirable to stay young and healthy as long as possible, to be as happy as possible, and to consume as much as possible. And the health systems in many countries are under increasing pressure to deliver on these goals.

Rising expectations affect not only the overall level of demand for health care, but also its composition. Patients not only want care, but the latest and best care, the fanciest technology, and the newest drugs. Knowing what is available elsewhere, they become skeptical (perhaps rightly) of the quality of care in their own local health centers or smaller hospitals. Consequently, they overwhelm regional centers and academic facilities. Even in many low-income nations, there is a growing urban middle class that increasingly inhabits a global world of information, expectations, and consumption. They pressure governments to provide expensive, high-technology care, from CAT scanners to organ transplants, and to spend a larger share of the national health care budget on a few elite institutions.

Political developments have reinforced these pressures. As nations have become increasingly democratic, it becomes increasingly difficult for political leaders to resist citizens' demands. When citizens are the subjects of a highly organized state, they are more likely to accept whatever health care a powerful government finds it expedient to provide. The active, self-seeking citizens of an open political order are less likely to be satisfied with such arrangements. Conversely, political leaders in a competitive system are more likely to use the health system to create tangible benefits for individual voters and organized interest groups, and hence to see health sector reform as an opportunity as well as a problem.

2.3. Force #3: Limited Capacity to Pay

Recent changes in the world economic order have created both problems and opportunities for many nations. But for many governments, the overall consequence has been a limit on the funds available for health care. Because of both their poverty and limited ability to generate domestic funds, few lower income countries are able to mobilize more than 1-2% of GDP for health, excluding international aid contributions. Lower transportation and communication costs have made it possible for nations to offer goods in other markets with unprecedented ease. Everything from cell phones and faxes to jet aircraft, laptop computers and the Internet now make genuinely global capitalism possible. Local sporting goods stores on five continents today sell Danish footballs (Umbro) made in Pakistan, Italian football shoes (Diadora) made in Bangladesh, and British shorts and shirts (Union Jack) produced in Guatemala.

At the same time, the sophistication of the economies in poor countries increases daily. Countries that used to make inexpensive radios now assemble advanced computers. Factories that once produced textiles now make the chips for those same computers. Brazil sells jet aircraft in the U.S., and U.S. computer companies open programming offices in Bangalore, India. News on all the world's stock markets can now be seen around the globe on the nightly business roundup from CNN.

The pace of change has been so fast that only older readers will remember that in the 1960s, transoceanic telephone calls were very expensive and had to be booked days in advance. Business travelers today can check e-mail and phone messages from a hotel room anywhere from Iguasu, Argentina to Chengdu, China, just as they can chat with colleagues anywhere in the world on a cell phone while driving down a highway that could be in any one of dozens of countries.

Worldwide, the second half of the twentieth century passed by much better than the first half. We have had our share of destructive wars and economic collapses, but nothing to match the two world wars and the great depression of the century's first fifty years. This relative peace and prosperity has given many individuals and corporations the opportunity and the funds to invest abroad. At the same time, national governments find their fiscal and monetary policy increasingly constrained by international capital flows and financial institutions. The international banks will not buy your government's bonds if you try to pay for social and health services by running too large a deficit. Nor will they invest in private sector projects if they fear fiscal instability, interest rate fluctuations, or swings in currency values. And loans from the IMF or the World Bank require a country to reduce its deficits, control public sector spending, and repay on time.

The emerging world economy has produced real economic growth in many countries. And this would seem to mean more available funds for healthcare. But often costs and demand in the health sector have grown even faster. And worldwide, real growth is extremely uneven. Until the mid-1990s, some Asian countries routinely achieved economic growth rates greater than 5% annually, while many African countries experienced negative or marginally positive growth. Moreover, the current economic order holds many dangers, especially for less advanced economies. A collapse in commodity prices can create serious fiscal problems for many governments. International economic linkages mean that economic difficulties in one country spread rapidly. When coffee prices fall, deficits in Kenya and Ethiopia skyrocket. If Brazil devalues, Argentinean exports to that country suffer. If the property market collapses in Kuala Lumpur, then Malaysia imports less, and its industries cut their prices to try to export more – actions that can cause serious economic problems for Malaysia's neighbors. If one currency falls, international speculative pressure shifts to another.

The combination of international competition and economic instability has produced various kinds of health care crises. In the first place, international competition makes businesses more cost conscious, in order to maintain export markets and prevent import competition. No one in the private sector wants to pay higher payroll taxes for health or social welfare purposes. At the same time, citizens complain that they cannot

afford to pay for increasing health care costs out of their own pockets. They also resist tax increases and/or higher co-payments. If these go up, how are they going to be able to afford the rising standard of living to which they aspire?

Governments find their fiscal capacities strained and limited. First, for most lower income nations, their ability to raise funds domestically is very limited. Low-income countries raise only about 20% of (a much lower) GDP in taxes, compared with 42% in high income countries (Schieber and Maeda, 1997). With public resistance to higher taxes, how can governments pay more for health care? Citizens also want better roads, housing, and education. Infrastructure investments are needed to support economic development. Meanwhile, multinational corporations demand tax concessions and threaten to relocate production elsewhere. And domestic firms claim the need for protection against international competitors.

These economic processes intersect with political pressures. Both the tax base and tax rate determine a nation's fiscal capacity, and these are determined by legislators and finance ministries. Here again, we see signs of change. As individualism gains in many countries, social solidarity has declined as a political force. Mobility and migration (intra- and internationally) weaken social bonds. Nations become more culturally heterogeneous, and the rich feel less obligated to the poor.

At the same time, providers inside the health care system struggle to preserve their incomes and protect their interests. In Eastern Europe, the large health care delivery systems created under communism are increasingly unsustainable, as governments work to restart economic growth and regain international competitiveness. As they try to control expenditures and prevent inflation, governments are tempted to hold down health care spending by limiting compensation to doctors, nurses, and other health personnel. These groups, in turn, work desperately to increase their falling incomes.

In other countries, politicians have made election promises they cannot keep with regard to various social programs, including pension funds. As those programs either go bankrupt or devour extra tax revenues, all other spending (including health) comes under pressure. In response health sector advocates often seek to remove health care financing from the general budget (e.g., by creating new social insurance programs) in order to give it protected status.

These political developments and economic constraints occur in the context of major social and intellectual changes, inside and outside the health system. These broad changes represent the fourth force that is driving health reform.

2.4. Force #4: Skepticism of Conventional Ideas

The global turn to the market has brought a trend toward diminished social solidarity and a broader turn against government in many countries. Often, the critique of

government is grounded in hard experience. In Eastern Europe and the former Soviet Union, the old model of centrally planned, state-owned enterprise has been widely discredited (even as older citizens remember the security it provided with nostalgia and regret). In many low-income countries bad governance has produced a squandering of limited resources, through poor policies and outright corruption.

Since the time of Ronald Reagan and Margaret Thatcher, enthusiastic advocates of an anti-government message have worked aggressively to promote their views. Private foundations, consultants, entrepreneurs, government officials, and experts from international agencies have all participated. In the health sector, political leaders have found themselves subject to the argument that the pro-market approach is the solution to all their problems.

These new ideas have also found acceptance because the public sector often performs poorly. In many countries, crumbling facilities, unhelpful staff, and low quality service are all-too-frequently characteristic of the public health care delivery system. This has spurred interest in "reinventing government," as well as new forms of corporatization and privatization throughout the world.

This pattern extends well beyond the health sector. Poorly run government bureaucracies and inefficient state-owned companies have proliferated because of perverse political incentives. It is attractive for governments to use the public sector to provide patronage rather than service, and to stress employment rather than output, in order to build up its political base. Workers are identifiable, easy to organize, and often grateful to political leaders for their employment; customers, on the other hand, are harder to identify and organize, and are likely to have less at stake. These factors lead to overstaffed, poorly managed, high-cost agencies. Until a few decades ago, urban political "machines" in the United States worked on the same principle, and the practice continues today in some public hospitals in New York, Boston, and Chicago.

But with countries increasingly short of money, there is rising criticism of such arrangements, both within the health sector and more generally. To break the politics-to-employment link, many institutional reforms have been attempted, ranging from the creation of quasi-autonomous public boards that supervise government activities, to outright privatization. These ideas pervade the debates over health reform throughout the world today.

But new arrangements will not necessarily perform better than current practice. As we argue below, the devil is in the details – especially when it comes to how organizations are managed and governed. To produce effective reform requires careful analysis, not simple slogans.

This trend against government includes a growing enthusiasm for the use of competitive markets to allocate resources. Competitive markets involve two key components: allowing consumers to choose, and making sure they have multiple options. In fact privatization alone is neither necessary nor sufficient to create such competition.

You can have private monopolies and competing public sector providers, but the two ideas are not often linked in practice.

In our view the applicability of the market model varies with specific circumstances. Is a given market large enough to support a reasonable number of efficient, independent competitors? Do customers know enough, or can they be given enough information, to choose intelligently? Will the savings that result from the competitive pressures on organizations to be more efficient be greater than the resources used up through higher transaction costs (i.e., the costs of arranging all the buying and selling)? These are empirical questions that a health reformer should answer before accepting the new intellectual currents. Furthermore, the market-based model has important ethical implications—as we discuss repeatedly, especially in the book's second half on the five control knobs.

3. Summary

Let us summarize the main points of this introductory chapter:

- Economic, social, and political forces, both inside and outside the health care system, are causing widespread pressures to spend more on health care.

- Furthermore, new technology (including new drugs), as well as changes in population structure and disease patterns, are continuing to push up per capita health care costs.

- The evolving international economic system often limits the capacity of governments and the private sector to meet these higher costs, whether through short-term crises or longer-term competitive pressures.

- Broad shifts in intellectual beliefs about the state and the market have led to widespread promotion of market competition and privatization as solutions for all economic problems, including policies for the health sector.

Many countries today face a gap between what they can pay for and what they would like to provide in the health sector. Expectations continue to rise as economies improve, countries become more democratic, and media-based images diffuse around the globe. In almost all countries, health care costs are increasing, due to changing demographics, evolving disease patterns, and new technology. And the implications of these changes are fought out in an increasingly diverse, open, and egalitarian set of political and social processes, in countries with severe budget deficits and limited economic resources for health.

At the same time, there is a global move toward new ideas for health reform, combined with an emerging recognition that getting health reform right is a complex social phenomenon indeed. However complex, we believe that the ideas in this book can

help analysts and decision-makers understand and manage the swirling processes of health reform. We begin our analysis, in the next chapter, with the health reform cycle.

References

Huber, 1999.

National Research Council, 2001.

Schieber and Maeda, 1997.

WHO, 1996.

Chapter 2
The Health Reform Cycle

1. Introduction

In Chapter 1 we introduced the idea that health sector reform can be viewed as a cycle. Now we want to analyze that process in more detail. In idealized form, the process of policy change moves through a cycle of six stages – problems are defined, a causal diagnosis is made, plans developed, a decision made, reforms implemented, and their consequences evaluated. Then, the cycle begins again, as new problems arise that must be addressed by policymakers. (Figure 2.1)

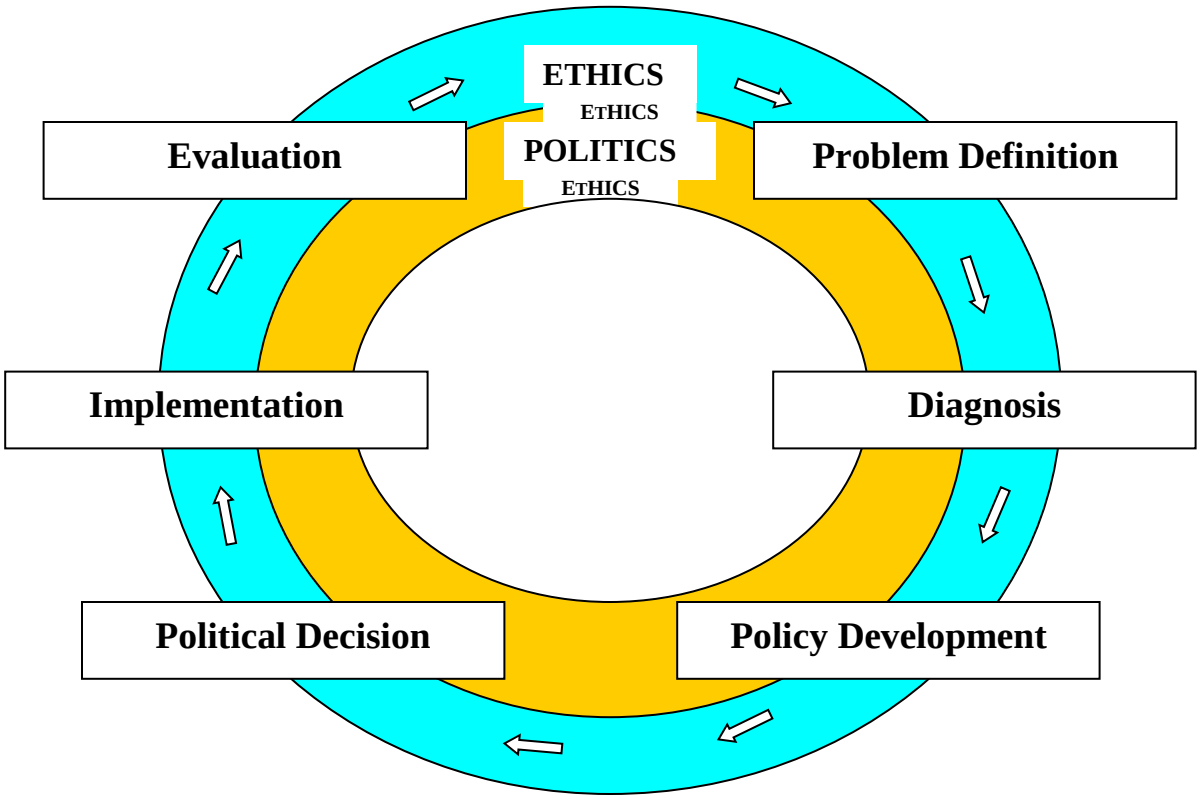
Of course, in the real world, reform does not always take place in such a simple fashion Instead reform efforts can begin in different places, skip stages, or several stages may occur at the same time. For example, health reformers might begin with a *diagnosis*, declaring, "The cause of all our problems is that public clinics and hospitals do not have enough resources to provide good service." Or a reform advocate might start with his favorite *solution*—"What our country needs is a new national social insurance scheme." Yet however the reform process actually unfolds the policy cycle framework provides a useful way of systematically analyzing that experience and of reminding practitioners of the various tasks they need to pay attention to at each stage of the reform process.

We have designed Figure 2.1 to call attention to some major themes of this book. In particular, the "ethics" and "politics" are there to remind us that decision-makers confront ethical and political issues throughout the reform cycle. Questions like "What is the right decision?" and "How can political pressures be managed?" pervade efforts at health sector reform. Our approach calls for an explicit examination of the ethical basis of health policies and rejects the notion that the values of all health systems are, or should be, the same. Because we believe that all policy positions (including our own) involve ethical dimensions, this book seeks to make our own values and goals transparent. Readers can then determine to what extent this perspective is useful to them. We also reject the view that the health sector reform process is exclusively technical. Instead, we call for extensive consideration of political realities at all stages of the reform process.

While six stages are identified in Figure 2.1, our work focuses most on the first four: problem definition, diagnosis, policy development, and political decision. This chapter, however, presents an overview of all six stages, and shows how these relate to our approach. The first four stages are then addressed in more detail in subsequent chapters – with each of the "control knobs" that are critical to diagnosis and policy development getting a chapter of their own in Part II of the volume. Moreover while we do not have a separate chapter on implementation, we have tried to provide practical guidance on that point in each of the control knob chapters, based on past experience from around the world. In our view, purely theoretical discussions of health reform options are not of much value. To actually move health reform forward reformers must realistically consider the difficulties of implementing a policy idea in their particular national context. And we try to help them do just that throughout the relevant chapters.

Finally, the cyclical nature of Figure 2.1 emphasizes a point already raised in Chapter 1. The problems of any health care system are seldom if ever resolved once and for all! As countries evolve, their health care systems have to respond to new challenges. Moreover successful reform may raise popular expectations, and thereby raise demands for further reform. But the process of reform is also imperfect. Reformers often encounter unintended consequences or discover defects in their plans that were not apparent initially. For any or all of these reasons, the cycle of reform will occur again and again. So let us now look at each stage of the health reform cycle in more detail and at the critical tasks each requires.

Figure 2.1: The Policy Cycle



2. Steps in the Policy Cycle

2.1. Step #1: Problem Definition

The most overlooked yet one of the most important steps in health sector reform involves defining the problem. Health care systems give rise to hundreds of statistics related to their performance. But which is an appropriate focus for public attention? When is performance "problematic" and thus an appropriate target for reform? As we discussed in Chapter 1, some sort of shock or crisis often initiates the reform process. But even then, contending forces and interests will have different views about how to define the problem and about what would constitute an appropriate solution.

Consider the following hypothetical example. Imagine a middle income country suffering from a region-wide economic downturn. The Ministry of Finance argues that the health sector is eating up too much of the governments' limited budget even as the medical society argues that "the problem" is that quality of health care is suffering because physician's salaries are not keeping up with inflation. Then the managers of the social insurance fund claim that "the problem" is the failure of the government to pay its share of the premiums for retirees, even as the association of hospital directors argues that the failure to provide capital for new technology "is the real problem". Clearly whatever group succeeds in having its problem definition be the basis for further discussion and policy development will have a great effect on what solutions are pursued and remedies accepted.

In thinking about problem definition we can ask two different kinds of questions. The first are *empirical* or descriptive: What social processes shape how problems are perceived? What factors determine the problem definition that reformers confront? We discuss how problem definition happens in practice in Chapter 5, on political analysis. There we focus on how to manage the political dimensions of health sector reform.

The second kind of question is *normative* or prescriptive: What makes for a *good* problem definition? What makes one area of poor performance a more appropriate priority than another? Answering this question involves ethics and philosophy. We introduce this topic briefly below, and then provide a more detailed discussion of the ethical aspects of health sector reform efforts in Chapters 3 and 4.

As stated previously, our normative approach to problem definition is based on the view that the health sector should be seen as a *means* not an *end*, that problems should be defined in terms of *outcomes*. As explained in the next chapter, different ethical theories do offer different perspectives on which outcomes matter most. Nonetheless, in Chapter 4 we will argue that it is possible to develop a set of performance criteria that encompass the major ethical perspectives that are relevant to health sector reform. One central concern we will suggest should be the *health status* of the population: How long do people live, what disabilities do they have, and how do these vary across various population groups? A second set of outcomes will highlight involves the degree of *satisfaction* that health care produces among the citizens of a society. A third set of outcomes we will urge require special attention involves *financial risk protection*. This is

what is the capacity of a health care system to protect individuals against the serious financial burden that can result from ill health. We will explain and defend these suggestions in detail below.

The focus on consequences that we urge is not an easy change for many reformers to make. Advocates with strong commitments to a particular policy often begin with their favorite solution. As the saying goes, “To a man with a hammer, everything looks like a nail.” They urge the nation to adopt competitive health insurance funds, or more family medicine or autonomous hospitals, without explaining how that would improve the performance of the health system. Moreover many in government are also not used to being performance oriented. They focus instead on spending budgets, meeting production targets or following rules. For them problems are defined in terms of the failure to fulfill these norms or rules, not in terms of the consequences of what is done or not done.

We understand from our experience around the world that old habits of mind die hard. As the historian of science T.S. Kuhn has explained, basic framing assumptions, which he called "paradigms", generally shape our thinking about any particular problem. And our emphasis on explicit means-ends analysis and critical strategic thinking represents something of a "paradigm shift" – a new way of thinking about health sector reform. But however awkward it may seem at first, we are convinced of the value of the intellectual and analytical discipline, to which this perspective leads. It forces reformers to identify goals and link proposed reforms to those goals, and thus to be explicit and self critical in ways that make it more likely they will achieve their objectives, whatever these may be.

This perspective also allows us to clarify the role of data in the process of defining problems and setting priorities, s___ health sector planners and professionals, seem to argue that all one needs to identify problems is good data. However as a descriptive matter, this is just not how the world operates. Many reforms move ahead without good data. Many well-documented problems are ignored in reform programs. For example, the financial hardships suffered by Chinese peasants who get seriously ill are well known, but this has not been enough to mobilize that nation’s policymakers.

Moreover from a normative point of view, data alone cannot completely define problems and priorities for health reform. Any policy decision, implicitly or explicitly, must rely on both science and ethics. For example, in many countries women live longer than men. Whether that represents an unfairness public policy should seek to correct cannot be settled by data alone.

Yet data do have a vital role to play. Yes information alone cannot reconcile conflicting positions based on genuine value differences. But scientific understanding can define options and clarify consequences. It can thus help move the debate to a more honest and consensual understanding of the choices available. For that reason, we believe that experts have a particular responsibility for helping the political process understand what is at stake in health reform. Otherwise, myth, ideology, and group interest will dominate the debate.

One important use of data in the process of problem definition is through a process called *benchmarking*, a term we borrow from the quality management literature. In an industrial context, managers use benchmarking when they look at the costs, defect rates, or productivity of competitive businesses to see what levels of performance they can reasonably aspire to. In health sector reform, benchmarking means looking at countries similar to one's own in income and spending levels, whose health system performance is particularly effective. Thus, reformers in Thailand might wonder why Sri Lanka has longer life expectancy while spending less on health care, and use that fact to focus their own problem definition. Similarly, Latin American countries could look at health statistics from Cuba or Costa Rica for setting their own aspiration levels. Despite the differences among countries, which we have repeatedly emphasized, international benchmarks can serve as a useful starting point for a discussion of performance problems. (There are other forms of benchmarking as well, and we discuss all of these more extensively at the end of Chapter 4).

2.2. Step #2: Diagnosing the Causes of Health Sector Problems

Just as a physician proceeds from symptoms to causes, a health sector reformer also has to undertake a *diagnostic journey* (see Chapter 6). Starting with a definition of problems based on health system outcomes, the next task is to work backwards to identify the determinants of unsatisfactory results. Physicians explore anatomy and physiology. Health sector reformers have to examine the causes of problems, by exploring what we call the *five control knobs of the health sector*.

This is not a simple task. One prominent Japanese quality improvement expert argues that the diagnostic process requires someone to "Ask why five times," to look for causes behind, beyond, and beneath the obvious. The task is to dig deeper and deeper in order to understand why the system behaves as it does.

For example, researchers interested in explaining the problem of higher than expected costs of health care in urban areas in China discovered very high use of certain pharmaceuticals. Going on a diagnostic journey revealed the following:

- In response to budget pressures, the Chinese government substantially decreased payments to hospitals;
- The hospitals were then forced to rely either on payments from the insurance funds that cover government and industrial workers, or on patient payments;
- The fee schedule for insurance reimbursement made high tech imaging and dispensing the newest pharmaceuticals particularly profitable;

- Based on traditional Chinese herbal medicine practice, the hospitals directly sold the pharmaceuticals their doctors prescribed. They also provided much of the country's outpatient care;

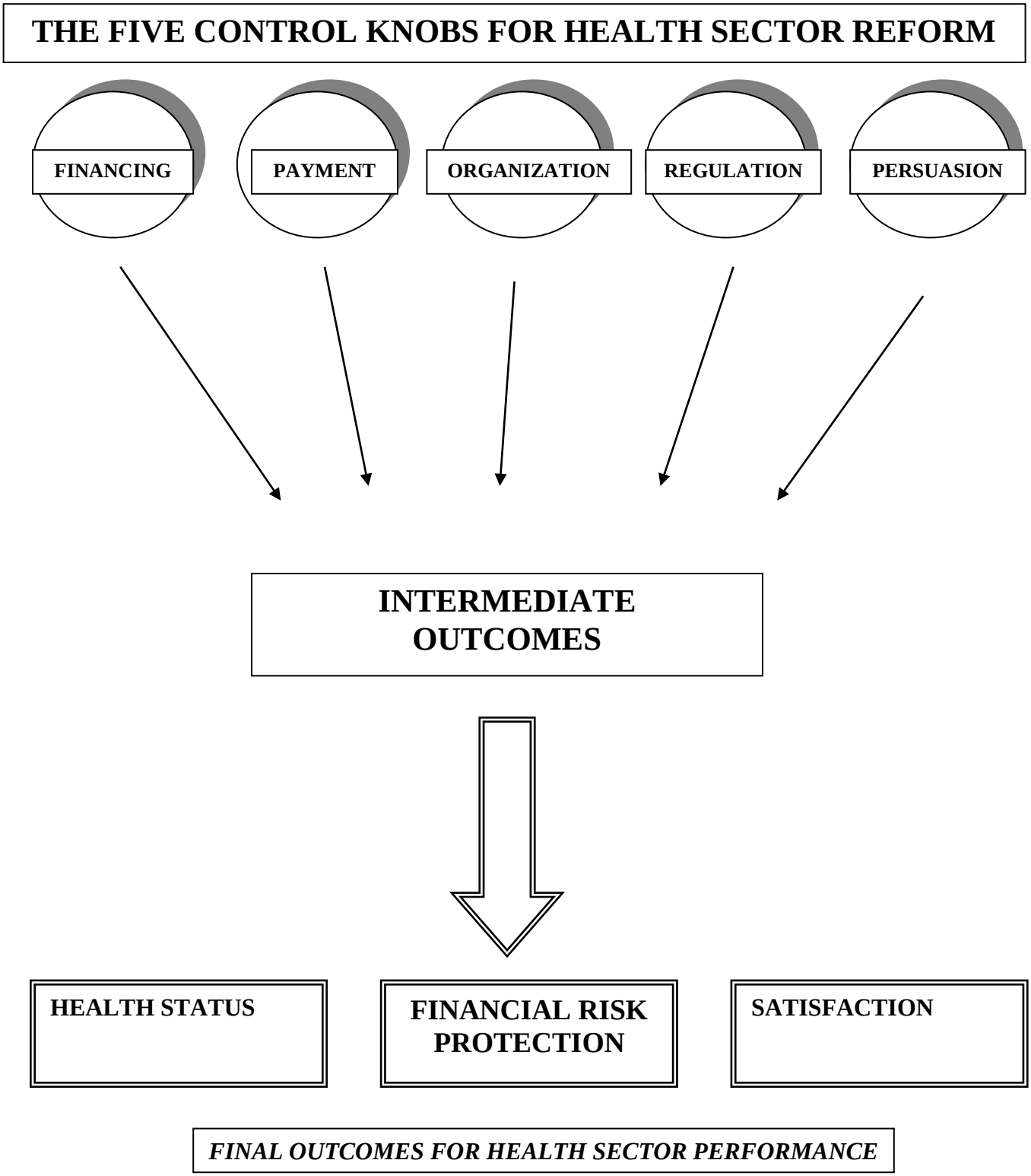
- To influence their physicians, hospitals began compensating them based on the profits to the hospital that the physicians' prescriptions produced.

As a result, prescription rates for the most profitable pharmaceuticals rose dramatically. Understanding this process does not by itself determine the "treatment" to select for this "illness" – one could change hospital funding, the insurance system, the price list, physician compensation schemes, or drug regulations. But diagnosing the causes of the problem is a necessary step toward devising an effective response.

2.2.1.The Five Control Knobs

To facilitate the diagnostic and policy development processes of health sector reform, as already mentioned we refer throughout this book to five "control knobs." This set of categories is designed to call attention to the variables that reformers can adjust to improve system performance. Thus not every important cause is part of one of the control knobs because it may well not be changeable by health sector reform efforts (Figure 2.2). While we examine them in detail in Part II, let us here introduce them briefly.

Figure 2.2
The Five Control Knobs for Health Sector Reform



Our categories are necessarily somewhat arbitrary. And other ways of talking about the system are possible. But we believe this five-fold framework is a useful way to sort through both causes and as a checklist to consider in devising solutions.

Financing: This refers to all mechanisms for raising the money that pays for activities in the health sector. It includes taxes, insurance premiums, and direct payments by patients. The design of the institutions that collect the money (e.g., insurance companies, social insurance funds) is also part of this control knob. (Chapter 7)

Payment: This refers to all the processes that provide money to health care providers (doctors, hospitals, public health workers, etc.) including fees, capitation, budgets, etc. These payment systems inevitably create important incentives to which providers respond. Money that comes directly from patients is of course a part of this control knob (Chapter 8).

Organization: This includes two sets of characteristics of the sector. The *macrostructure* describes how organizations divide up various activities, such as who does what, and who competes with whom. The *microstructure* describes how entities are organized internally – how tasks are divided and employees rewarded. The microstructure helps shape how institutions respond to the incentives that the payment system creates. (Chapter 9)

Regulation: This involves all the efforts by the state to change the behavior of those in the health system including providers, insurance companies and patients. We note that just because a regulation is on the books does not mean it is implemented and enforced and consider what it is a reformer has to bear in mind to insure that regulation works as intended. (Chapter 10)

Behavior: These are efforts to influence the behavior of individuals (both consumers/patients and individual providers). This control knob includes everything from mass media campaigns on smoking, to HIV prevention efforts, to attempts to use the Medical society to influence physician behavior, to the organization of local women's groups focused on family planning. (Chapter 11)

The control knob variables explain many aspects of health system performance. They influence both what organizations exist in the health system and how these organizations perform. The *financing system* determines what resources are available. The *payment system* then determines on what terms those resources are available to individual organizations. The *organization* of the health sector determines the kinds of organizations that exist and their internal structure - which in turn shapes the behaviors they tend to exhibit. The *regulatory system* imposes a set of constraints on those behaviors. Finally, *efforts to change behavior* influence how individuals respond to health sector organizations and in turn shape the opportunities that organizations confront.

However the control knob variables do not explain everything! Many other cultural and structural variables also influence the outcomes of the health sector. For example, understanding why a country's regulatory agencies do not function effectively may well require us to understand a nation's political institutions and cultural traditions – factors that may not be easily affected by health reformers. We will emphasize the role of the control knob variables exactly because they are controllable. However, delineating the role of not easily controllable variables, like culture, history, and tradition can help health reformers understand the problems they confront. Engineers designing automobiles can not change gravity, but they have to work with its effects in mind. We will return to this question at the beginning of Part II when introducing the control knobs, we will provide an overview of what they omit as well as what they include.

2.3. Step #3: Policy Development

Once the causes of a problem have been diagnosed, the next question is what to do. What is the right therapy for the problem? What is an appropriate policy to address the causes? The process of policy development is more difficult than it might seem. In this section, we first consider why new policy ideas are difficult to develop, and where such ideas are coming from. We want also to explain why it is important to look *forward* in the policy cycle to matters of politics, implementation, and evaluation when doing program design. Finally, we consider how the *process* of option development influences what happens at later stages of the policy cycle.

2.3.1. New Ideas

Human thinking is heavily ruled by assumption and habit. We have already mentioned how T.S. Kuhn, described the role of "paradigms" in shaping how questions are posed, what kind of evidence is acceptable, and what kinds of answers are viewed as legitimate. Professions typically educate their members in a particular paradigm and reinforce these ways of thinking through practice and experience. Economists, for example, tend to view everything in terms of prices and markets, just as public health specialists focus on epidemiology and prevention. When powerful organizations adopt and promote the dominant paradigm, change is even more difficult. For example, the World Health Organization promoted the concept of Primary Health Care for several decades, which significantly decreased the attention given to alternative policy approaches.

Where they can reformers look for new ideas? International learning is one helpful approach. For example, if your country wants to develop urban health clinics, it would be sensible to examine the experience in other countries, and the lessons from their failures or successes. Yes, economic circumstances and political institutions do vary so that ideas from other nations need to be adjusted to local conditions. But none-the-less the volume of experimentation around the world has been very substantial in recent years and is a useful place for a reformer to get started.

Policy innovations outside the health sector are another source of new ideas. In the United States, for example, the idea of vouchers to the poor, who then use them to

help purchase their own housing, was later extended to school finance, and is now being considered for health insurance. Similarly in our discussion of social marketing in Chapter 11, we review how ideas developed in a commercial context can have health sector applications.

A third source of ideas is not experience but theory. We have already expressed our skepticism about arguments not rooted in real life examples, Yet theory can be a useful source of provocation and inspiration – provided implementation issues are well considered. For example much of the argument for so-called "a____-markets" and "purchaser-provider separation" had their roots in economic theory.

2.3.2. Looking Forward

To develop an effective plan for health sector reform, reformers have to think ahead to political decisions and implementation – and design policy with those tasks in mind, (as we discuss more fully in Chapter 6).

The first kind of forward thinking involves politics (see Chapter 5). Sensible reformers begin to think about the political acceptability of their policies as they are developing them. Why devise an ideal plan that has little chance of being adopted?

Similar considerations apply to implementation. We have noted how countries vary in administrative capacity and in the attitudes that citizens have toward government. For example, Hungary began paying hospitals a fixed fee for each type of admission. To raise their reimbursement, however, many hospitals reported nearly all baby deliveries as complicated. A country where providers are likely to "game the system," therefore, must anticipate those responses in policy design. Similarly, many clinics in Thailand only establish medical records for seriously ill patients. Given manpower shortages, this may be a sensible use of their limited resources, but it means that certain kinds of regulatory options are not available to health reformers.

Another implementation issue is whether to proceed immediately to full-scale nationwide implementation, or to take a more limited demonstration project approach. We have noted that reform often occurs in the context of a crisis. In such situations, it is tempting to move directly to a bold initiative, while public attention is focused on the problem and supporting forces mobilized. On the other hand, an experimental or demonstration project approach allows for learning. Why not try out a new idea in a few cities or regions first to reveal difficulties in implementation and flaws in the plan's logic before it is put in place nation-wide?

Viewing policy innovation as an experiment also allows reform to be designed to evaluate competing proposals. For example, in recent health insurance reform efforts in China, first two, and then twenty municipalities were involved in various experimental programs, before the national scheme was finalized. On the other hand, experimental programs can attract energy and enthusiasm that can lead to a degree of success that is hard to emulate when a program is "scaled up" for system-wide adoption.

2.3.4. The Design Process

The task of developing reform plans is as much political as it is analytical. This refers both to the *content* of the reform ideas, and to the *process* used to develop these ideas. For this process can be a key step in mobilizing support for reform. Thus, the process of policy development should be designed to produce a plan that is both technically sound, and that has a good chance of acceptance in the political domain.

A classic example of how *not* to do policy design in this regard is the failed Clinton health reform (Skocpol, 1996). The plan was developed behind closed doors by technical experts, with no consultation with either the major interest groups or political leaders in Congress. As a result, even groups that stood to benefit from the reforms, such as hospitals that would have received additional revenue from previously uninsured patients, were slow to support the scheme and never did so enthusiastically. Similarly, legislative leaders who were excluded from the process never felt responsible for the plan. As a result, none of the committees in Congress (all controlled by Democrats) even produced draft legislation that embodied the main ideas of the Clinton plan.

Having diverse perspectives involved in policy development can serve both political and technical ends. Giving potential supporters a role in the process can help transform them into actual supporters. The process can then also provide a testing ground for new ideas, and help protect experts from becoming prisoners of their own enthusiasms. Involving those who actually work in the system can also result in substantive changes that improve the chances of effective implementation. Of course too much "participation" can limit reformer's ability to get what they want and need from the process. Thus as well as overseeing the actual day-to-day operation of the group, designing the membership terms of reference and staffing of whatever task force, committee or staff group is created to actually do detailed policy development, and doing this task well may be key to having a reform that succeeds in the long run requires a good deal of skill and experience.

2.4. Step #4: Political Decision

As shown in Figure 2.1, each stage in the policy cycle is affected politics, including problem definition policy development, and implementation. The adoption of a reform proposal, however, provides a focus for political decision-making, often by the executive and legislative branches of government. As Chapter 5 examines the politics of health reform in greater detail, we will only address these processes briefly in this section.

Health reform typically confronts difficult political challenges. Organized interest groups (physicians, hospital owners, and the pharmaceutical industry), with a large stake in the current system, are likely to oppose reform. On the other hand, the intended beneficiaries of health reform (sick patients, the poor, the underserved) are often less powerful and less well organized. Some potential beneficiaries within the health system (e.g., new family doctors yet to be trained) may not yet exist, and so cannot play an effective role. As always, in the pull and tug of political decision-making, the future tends to be underrepresented relative to the past.

Getting health sector reform adopted is not just a matter of commitment (political will); it is also a matter of effective political strategy and coalition building (political skill). Like other skills, political skills can be analyzed and understood, taught and learned – and for health reformers, political skills are essential. Whether a proposed reform is adopted will depend on the skill, commitment, and resources of its proponents, and on the political strategies they employ.

There is nothing dishonorable in being politically astute in the service of health sector reform. Politics is how governments do their work and make decisions. In democratic states electoral and legislative politics may be key, and one-party systems, politics – in the form of bureaucratic politics – plays a comparable role. Decisions are made through intense struggles within the state apparatus, as agencies and factions fight for control. Someone who truly cares about results in health reform thus must learn to think politically and act strategically, no matter what kind of system they operate within.

A key political task in all situations is building a strong support coalition. As we discuss in Chapter 5 this means identifying groups and individuals with enough collective political power to get the proposed program accepted. The political strategist for health reform needs to map the power and position of key groups, and to devise strategies that will produce the required support. These strategies also need to take into account the country's institutional and political structure. Moreover politics doesn't end once a plan is adopted – but instead can also play a significant role in whether or not a reform is conscientiously implemented – which brings us to the next stage in the cycle.

2.5. Step #5: Implementation

The American political scientist Aaron Wildavsky once said, "It is no use having a good idea if it cannot be implemented." Many allegedly good health reform ideas have failed because, in practice, they were *not* good ideas – exactly because they could not be implemented.

Health sector reform *always* requires organizations and individuals to behave differently. Without such changes, nothing new would be accomplished. Yet change is almost always resisted. One reason is psychological. New procedures and structures are strange and unfamiliar. For many, the mere fact of newness creates anxiety and resistance. The old adage, "Better the devil you know than the devil you don't," embodies this idea. Resistance to change also arises because change can bring real costs. New procedures and arrangements take time and effort to learn. Existing hierarchies can be upset. Those who benefited from the old system can lose greatly. Moreover, those most successful under the old system have the most to lose; hence, the currently powerful are often the most opposed to reform.

Another force that inhibits change is the difficulty we all have in giving up familiar ways of thinking. Ingrained patterns of thought and behavior, which have evolved slowly and served well in the past, can have a strong hold over us. Moreover

established patterns of interaction are often strongly supported by a network of mutual expectations. If a doctor wants to treat a nurse differently, or a patient wants to relate to a doctor differently, they can encounter all sorts of pressures to conform to expected behavior. To most of us, these assumptions – like many features of our culture – are invisible. Yet implementation of health reform can depend on changing some of these same cultural assumptions, a process that can be very difficult.

Overcoming these resistances to change requires sophisticated management leadership. Yet such leadership is often in short supply. Health sector reformers (including ministers of health or finance) are often physicians, politicians, or economists. Few have had experience as the chief executive of a large organization, and they rarely know the importance of management skills. Senior academic physicians can be particularly deficient in this regard, given the very hierarchical nature of academic medicine in many countries. They often do not realize that giving instructions is not sufficient; that reforming a health system differs from performing surgery in an operating room. Implementing health reform involves basic organizational tasks: assembling a team, developing a plan, assigning tasks, devising and coordinating schedules, and monitoring and motivating subordinates. Given that ministers often serve less than one year, and frequently less than two, perhaps it is not surprising that too often not enough attention is paid to implementation. But when that occurs, results are likely to be commensurably disappointing.

To be implemented effectively, health sector reform must be monitored and observed, so problems can be identified and corrective measures instituted. An appropriate monitoring and reporting system is therefore key to successful implementation. But many problems can arise in collecting the relevant information. The incentives on providers to misrepresent, even lie, can be substantial. The data that is easily available may not exactly measure the performance reformers care about – and that can lead to incentives that distort behavior. Thus a key task for reformers is to design a monitory system that balances costs against data quality, and ease of operation against resistance to fraud. For such a system is key not only to effective implementation, but also to the evaluation of health reform.

2.6. Step #6: Evaluation

Experienced evaluators know that successful evaluation has to begin well before any new program is actually implemented. Determining the effects of a new policy in a changing world is inherently difficult. The question of whether what happens is the result of reform, or would have occurred anyway, often arises. Since data gathering consumes scarce administrative and organizational resources, a series of strategic decisions must be made if the evaluation process is to be both effective and sustainable.

The simplest evaluation approach is a before-and-after comparison, this involves examining how some important outcome changes over time as the policy was implemented. But this approach does not reveal the effects of other events or circumstances. For example, improvements in health status could be due to a surge in

economic growth, rather than the introduction of a new financing or physician payment scheme.

The classic solution to this problem is to establish a "control group;" a population *not* subject to a new program or policy, but which is studied in the same way as the experimental group. This allows a "difference in differences" approach, which compares the before-and-after differences between the experimental and control groups. This approach works best in conjunction with a demonstration project strategy, since a nationwide reform leaves no room for a control group for comparison purposes.

When no control group is available, an evaluator has several fallback options. Perhaps different regions of the country vary in the extent to which potentially confounding variables influence outcomes. Alternatively, a similar neighboring country could provide something of a baseline. These issues need to be thought through carefully in designing the evaluation process. But even the simplest before-and-after comparison for a new policy cannot be made unless there are good "before" data. Health reformers, therefore, need to collect baseline data well before any new policy or program is instituted.

While we cannot explore the many relevant issues to evaluation in health reform in this book, health reformers should recognize a few key lessons about the role of data in evidence-based evaluation. First, data collection is not free, and better data typically cost more to collect. A good scheme for evaluation must therefore balance cost against usefulness. Second, the costs of data collection often fall on the people doing the reporting. If the costs to them of providing good data are too high, they will provide poor data. For example, in many states in the U.S., hospitals have to report the number of prenatal visits of each woman they deliver. But the medical records of the women are often incomplete, and obstetricians frequently simply fill in the ideal number, because they have no reason to take the time and effort to obtain more accurate information.

Third, it is possible to collect too much data. Over collection can produce "data cemeteries," in which piles of "dead" data accumulate, too massive for anyone to analyze. A key question for any manager to ask of a proposed data collection effort is, "How will I use these data once I have them?". Data that are not linked to actual decisions are likely to be interred in an early grave by those too busy to look at them.

One strategy for getting reliable information is to use data that organizations collect for their own purposes. Such data are more likely to be collected and recorded accurately since managers need them to run the organization. In some countries, payroll data, personnel records, or hospital discharge records may be relatively accurate. When such records exist, reports are potentially auditable against the underlying sources, which can increase reporting accuracy.

In sum, evaluation has to be an early concern for health reformers. It cannot wait until a new program is being implemented. Baseline data have to be collected before implementation, and administrative systems have to be designed with evaluation in mind.

Unintended consequences are frequent in health sector reform. To make a difference, health reformers must take this stage of the policy cycle seriously, so that such consequences can be detected and responded to. And when it is done well, the data collected will reveal new problems that start the cycle all over again.

3. Conclusion

This chapter has provided an introduction to the policy cycle for health sector reform. In subsequent chapters, we will explore selected aspects of the cycle in greater detail with attention to five main themes:

Step #1: Problem Definition:

Health systems are social responses to social problems. Problem definition and health reform priorities are always based on value choices. We believe these choices should be explicitly stated and examined. Our approach to these choices includes a guide to *ethical theory*, which we address in Chapter 3;

For sound policy development, problems should be defined in terms of *health system performance objectives*. Although health reformers often hold multiple objectives, we believe that a short list of core health system objectives – those which in our experience are the most relevant to health reform – can be identified. We present these objectives in Chapter 4, and link them to the ideas of ethical theory presented in Chapter 3.

Step #2: Diagnosis:

In order to identify potential solutions to improve performance, assessment of health system performance requires a diagnosis of the causes of poor performance. We have designed a method of diagnosis that is systematic, can be taught, and can be practiced. We elaborate this approach to health system diagnosis in Chapter 6. The approach is based on a causal model of the determinants of health system performance.

Step #3: Policy Development

Health systems are complex, with many steps that link broad policy strategies with final outcomes. We propose conceptual framework of five "control knobs" that include the causal variables that meet two essential criteria: they are significant determinants of health system performance, and they can be manipulated through policy reform and effective implementation. In Chapters 7 through 11, we provide detailed discussion of these. For each control knob, we provide practical guidance on design and implementation of interventions and on the impacts of different reform strategies using the control knobs.

Step #4: Political Decision

In health reform, politics matters throughout the policy cycle, and we have integrated our concern with politics in each chapter of this book. As presented in Chapter 5 we believe that systematic *political analysis* and the use of that analysis to develop a *political strategy* is essential for effective reform. The placement of political analysis in this book reflects our view that it should be done early and often throughout the policy cycle.

These five elements – the importance of ethical theory, explicitly identified health system performance objectives, a systematic approach to health system diagnosis, use of the health system control knobs, and political analysis – are the core of what is new and useful in this book. These points are developed in detail in subsequent chapters, where we provide the substance of our approach for getting health reform right.

References

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

Skocpol, Theda. 1996. *Boomerang: Clinton's Health Security Effort and the Turn against Government in U.S. Politics*. New York : W.W. Norton.

World Bank. 1993. *World Development Report: Investing in Health*. New York, NY: Oxford University Press.

Chapter 3
Judging Health Sector Performance: Ethical Theory

1. Introduction

Arguments about what society *ought* to do always involve ethics. This is so even when the person asserts that the argument is purely technical and does not involve values. Economists often claim that they are not concerned with ethics, only with economic efficiency. Public health experts often take similar positions, arguing that they do not do ethics, they only protect the public's health. But how do we know that efficiency or health are the most important goals that society should try to achieve? Suppose that efficiency conflicts with fairness or that efforts to improve health lead to restrictions on individual liberty? Why value health or efficiency over these other objectives?

This book is based on our firm belief that judging health sector performance requires ethical analysis. In this chapter, we introduce three major ethical perspectives as a basis for making such judgments: *consequentialism, liberalism,* and *communitarianism.* Our purpose throughout is practical. We explore moral philosophy not for its own sake, but rather as a tool for making real-world judgments. We believe that understanding the larger ethical perspectives that lie behind debates on health reform can help policy analysts and policymakers do their jobs more effectively. They can better explain and defend their own positions, and can better understand and respond to the positions of others.

The first doctrine we discuss, *consequentialism,* says that we should judge a policy by its results. The most popular form of consequentialism is *utilitarianism*, which evaluates consequences by examining the well-being of all individuals in a society. This perspective motivates many health reform efforts around the world.

But health reformers care about more than consequences. They are often concerned about how the health system relates to the distribution of individual rights and opportunities. This doctrine, known as *liberalism,* constrains the means we use to pursue consequences through the protection of individual rights. Liberalism also can imply that some of us have obligations to ensure the rights of others.

The third perspective, *communitarianism*, states that neither consequences nor rights are the appropriate focus of concern. Instead, what matters is the kind of society we have and the kind of individuals who live in it. In this view, communities have an obligation to raise their members to share the community's notions of virtue and good behavior. Communitarianism can conflict with both consequence-based and rights-based thinking, since inculcating virtue can constrain liberty as well as reduce efficiency.

In this chapter, we discuss these three ethical theories and explore their strengths and limits. Each theory has distinct positions under the same broad heading, and we present two distinctions for each theory. In the following chapter, we will show how

these theories relate to the most common criteria used in health reform and illustrate how ethical analysis can help clarify the objectives and processes of health reform.

2. Ethical Theory #1: Consequentialism

The most obvious approach for making ethical judgments about the performance of the health sector is based on consequences. This line of reasoning argues that "The end justifies the means" and seeks to maximize positive consequences. This focus on ends is an area that critics of consequentialism often address: Which ends should be pursued? Which consequences count? How can consequences be measured in a practical (and defensible) way for the purpose of health reform?

There are two major intellectual traditions about how to judge consequences of social policies: subjective utilitarianism and objective utilitarianism.

2.1. Subjective Utilitarianism

The first position derives from the work of the nineteenth century English philosopher, Jeremy Bentham, who believed that individuals could best judge for themselves what makes them happy. He argued that people experience different levels of utility in various situations, depending on their particular tastes and preferences, and his doctrine came to be known as utilitarianism. He proposed that the rightness of an action was not intrinsic, but was determined by the "hedonic calculus" of adding up the pleasure and pain it produced. The right action is the one that produces the "greatest happiness of the greatest number" (Bentham, 1789). According to this approach, the way to evaluate the performance of the health care system is to ask everyone how happy it makes them, or what level of utility the system produces.

In certain ways, utilitarianism is a radical egalitarian concept. It posits that all individuals matter and that all matter equally. The task of judging a policy simply requires us to add up everyone's utility level. It also decentralizes evaluation, since all individuals judge their own happiness for themselves. Bentham argued that all tastes were equally valid and utility was a matter of each person's subjective experience. We call Bentham's position *subjective utilitarianism*.

Economists who use this analytical framework take the process one step further. They propose to use *cost-benefit analysis* to determine which action produces the greatest utility. This involves asking all potential beneficiaries of a policy or program to evaluate those benefits in terms of their own personal *willingness to pay* for the expected gains. That willingness is then compared to the cost of the program. If what people would be willing to pay is greater than the cost, then the policy would advance the greatest good for the greatest number and should therefore be adopted.

Economists want to compare benefits with costs because the cost of a program actually represents the benefits we give up elsewhere in the economy when we

implement a program. These foregone benefits are called *opportunity costs*. Economists believe that in a competitive economy, the price for an input—like labor, land, or capital—will be set by the value of the outputs that the input can produce. A businessperson will pay a worker $10 an hour, if, in an hour, the worker can produce goods that can be sold for more than $10. If a program costs $100,000, then the inputs used in the program could have produced goods worth $100,000 if they had been used elsewhere in the economy. Thus, cost-benefit analysis is really benefit-benefit analysis. Subjective utilitarians tell us to do those activities where the benefits we gain are greater than those we give up.

Evaluating health reform through formal cost-benefit analysis is not easy. One reason is that it is difficult to know what people would be willing to pay for services they do not yet have. Will they answer questionnaires honestly? Do they even know what they would pay for a hypothetical benefit?

Nevertheless, subjective utilitarianism has been very influential in discussions about health reform. Policymakers are often urged (especially by economists) to use markets to allocate health care. With markets, analysts do not have to guess what health care consumers would be willing to pay, since they can observe what consumers actually do pay. With markets, purchasers get what they are willing to pay for, and consumers buy those goods and services where their willingness to pay exceeds the costs of production. The influence of market-based policies in health reform derives partly from the subjective utilitarian emphasis on seeking to maximize social welfare through this focus on individual preferences.

Critics of subjective utilitarianism argue that willingness to pay depends on a person's income and is therefore biased against the poor—you cannot be willing to pay more than you have. Subjective utilitarianism can lead to the conclusion that we should not provide costly services to the poor, since they would be unable (and hence unwilling) to pay for them. A true subjective utilitarian would argue that if we want to make the poor as well off as possible, then we should give money to the poor and let them spend it as they choose. If other things are more valuable to the poor than health care, that should be their choice.

This line of reasoning implies that health is not a special good, but is just one more thing citizens buy or not, as they choose. It also implies that such choices should be respected, and that citizens know enough about what will make them happy to make these decisions. Not everyone agrees on these assumptions, leading us to the objective utilitarian approach.

2.2. Objective Utilitarianism

What if you believe that health reform should be judged by its consequences, but do not think that individuals make good choices? For example, should governments assure that markets provide tobacco smokers with what they want? If doctors significantly influence

how patients are treated, is the free market really the best model for thinking about the health sector?

People who want to promote individual well-being but are skeptical of how well the average citizen can make choices typically conclude that a group of experts should define well-being in objective terms. They urge us to develop an index or measurement that can be used to evaluate everyone's well-being, rather than determine well-being by adding up individual preferences. We call this position *objective utilitarianism*.

Objective utilitarianism centralizes what Bentham sought to decentralize. The single index of well-being disregards individual variations in preferences. One individual may want to avoid physical disability or chronic pain, even if this means a shorter life span; another may seek the longest possible life, whatever suffering that life entails. But such variations in perspective have no place in an objective utilitarian calculation based on a single index for evaluating consequences.

Objective utilitarianism has a long history in public health – illustrated by Florence Nightingale's efforts to show it was more cost effective for the British Army to care for the wounded than to let them die and then train new recruits. There were various attempts to construct health status indices well before World War II. Objective utilitarian approaches have also been employed in clinical research where measures of "quality of life" have been constructed to evaluate the results of alternative treatments. It is also the philosophical position behind the analysis of disease burdens through measures such as DALYs (Disability Adjusted Life Years) or QALYs (Quality Adjusted Life Years), as conducted by the World Bank (1993) and the WHO (2000). This approach tells us to pick policies with the "biggest bang for the buck," with "bang" measured in terms of health gains. One example was the Oregon Health Plan in the U.S., proposed in 1994, which was intended to rank all medical treatments covered by the Medicaid program (for low-income individuals) in terms of QALYs per dollar, and restricted coverage to the most cost-effective interventions (Bodenheimer, 1997).

When implemented, the objective utilitarian approach becomes *cost-effectiveness analysis*. This approach has been widely used outside of health care, under various labels. During World War II, British mathematicians used such methods to calculate the best way to conduct military operations (e.g., searching for a flyer downed in the English Channel) giving rise to the term *operations research*. In the 1950s, similar techniques were applied at the Rand Corporation in the U.S. to evaluate the comparative efficiency of different weapons systems, in what was called *systems analysis*.

The measurement of health gains confronts many technical problems. One critical question is how to combine extensions in the length of life with improvements in the quality of life. The QALYs and DALYs methods use a particular (and controversial) solution. A year of life in good health counts as one unit of gain. A year while suffering some disability counts as a fraction of this, say seven-tenths of a unit. The value of life extension is then counted as the quality of life of the person saved, multiplied by the number of years provided. An improvement in quality is measured by the fractional gain

in quality in any one year, times the number of years it persists. Thus, adding 5 years of life to a person at .8 quality level is calculated as:

(8 X .5) = 4 QALYs

Similarly, raising someone from a quality level of .6 to a level of .8 for 20 years does the same:

(.8 - .6)20 = 4 QALYs

The controversy arises over whether these two amounts of health gain really are equivalent.

The construction of a health index thus necessarily involves values. For example, how should mental disability be counted in comparison to physical disability? Are years of life lost at different ages to be counted equally or unequally? The DALY index, for example, values years of life in middle age more highly (Murray, 1994). This reflects a Western, productivity-oriented perspective, but could be controversial in a society where the old are viewed as a source of special wisdom.

A health status index thus embodies a particular viewpoint about the relative desirability of different kinds of health gains. A major issue is whether there is a single transcultural basis for health measurement, or whether culture inevitably enters into all measures of health status. The advocates of such health measures must contend implicitly or explicitly with the idea that not all societies value health outcomes in the same way. In practice, they have convened panels of experts and used consensus-building methods to support the selection of a single set of weights for evaluating different kinds of health gains and losses all around the world in the same way (Murray, 1994).

Some health policy analyses do not use comprehensive measures of health gain. For example, many health ministries have used measures like the infant mortality rate (IMR) for policymaking, while in the U.S., environmental policy has been analyzed in terms of mortality, the expected number of lives saved. These more limited indexes require less data than comprehensive scales like DALYs. Their disadvantage is that, because they are not comprehensive, they can give a distorted picture of the situation. For example, the development of a neonatal intensive care unit in a community hospital can have the effect of *increasing* the observed IMR, as sick newborns live long enough to be counted as neonatal deaths rather than as stillbirths. Thus, any partial index needs to be employed carefully.

Another problem with consequentialism is the assessment of causation. To use consequence-based analysis, one must evaluate the consequences of past policies and predict the consequences of future policies. This can be quite difficult. In some countries, for example, differences in the reported IMR reflect variations in local customs or bureaucratic competence as much as variations in actual health conditions. Similarly, to select policies based on their expected impact, we need to be reasonably confident about

our predictions of a policy's likely consequences. Making good predictions can be complicated. Suppose a country is considering a proposal to set regional caps on physicians' total earnings. If physicians are sufficiently powerful to overturn such limits within the administrative process, it would be foolish to base a decision on the assumption that the policy will be implemented as planned.

2.3. Some Utilitarian Complications: Uncertainty and Time

Two technical problems with utilitarian analysis deserve additional consideration: uncertainty and time. The consequences of health reform are seldom predictable with great accuracy. Even if the future were relatively certain, we would still have the problem of evaluating consequences that may occur in the future. We briefly discuss these two problems in counting costs and benefits, and refer the reader to other references on cost-benefit and cost-effectiveness analysis for more detailed treatment (Weinstein and Stason, 1977; Wenz, 1986).

Let us consider uncertainty first. For a subjective utilitarian, the way to value uncertain gains, in theory, is straightforward. Uncertain benefits and costs are evaluated by asking the affected person how much the uncertainty changes the value of the result in question. Uncertainty is incorporated into direct questions about a person's willingness to pay.

This response, however, introduces a complex set of assumptions and implications. In particular, it requires us to assume that individuals have adequate knowledge about their own utility levels: to weigh uncertain gains and losses, people have to know not just whether they prefer outcome A to outcome B, but also *how much more* they value one over the other. Whether individuals have consistent and stable attitudes towards such choices has been a subject of substantial debate and empirical research.

The problem of time raises similar issues. Here the debate focuses on the discount rate. In financial markets, returns this year are, in effect, worth more than returns next year. If the interest rate is 7%, it implies that next year, you have to give me 1.07 times what I lent you today. This means that I value dollars next year *less* than dollars this year, since you have to give me more of them. This process continues for gains that occur further in the future. The procedure for computing the decrease in value of future gains is called *discounting*, and the interest rate used in the calculations is called the *discount rate*.

When evaluating health reform, should future health benefits be discounted? And if so, at what rate? Some analysts argue that, as a matter of social policy, future gains should be valued the same as current ones, since discounting effectively discriminates against the future. Others believe that if we don't use discounting, we then encounter an investment paradox, as follows. Suppose we take $100,000 from current health spending and put it in the bank. Since science is always advancing, and we can earn interest, that $100,000 will produce more health gain in 10 years than it does today. Thus, without discounting, we would defer far too much spending to the future.

In addition, the discount rate selected for analysis can have a major impact on the relative attractiveness of policies that produce short-run versus long-run gains. At a 7% discount rate, gains 10 years from now are worth half of what they are today—and gains 40 years from now are worth less than 6% of their face value. Lower discount rates give greater relative value to future benefits (long-run gains), and will be more favorable to projects that promise long-run gains over those that promise more short-run gains. There are two approaches commonly used for setting the discount rate.

Subjective utilitarians argue that rational individuals will adjust to the opportunities offered by the market to borrow and lend. This allows individuals to shift their own gains to the present (by borrowing) or the future (by lending or investing). They conclude that the right discount rate to use is the market rate that individuals actually confront, suitably adjusted for differences in risk and the rate of inflation. According to this argument, if individuals are the appropriate source of values, then we should use a discounting method that reflects individual valuations.

In contrast, some utilitarians are skeptical of how well capital markets function in practice. They believe that society should make an independent valuation of gains that occur at different times and argue for a *social discount rate* (which represents more of an objective utilitarian position) instead of a *market discount rate*. This debate over the discount rate reflects a major problem for utilitarians, because changing the discount rate can change which policy is ranked highest by economic analysis. Every numerical analysis of future gains and costs must use some method to deal with the time problem, thereby introducing important value judgments into the foundation of any utilitarian calculation.

A final problem for health reformers who consider the utilitarian approach is the ruthlessness implicit in the rule: Greatest good for the greatest number. This perspective tells us to spend our limited resources where they do the most good, regardless of the fairness or equity of the result, and regardless of the negative consequences for particular individuals or specific groups, as long as the total results are beneficial in the aggregate. Utilitarian policymakers can disregard patients who are too expensive to save, and can sacrifice the few for the sake of the many. Following utilitarian logic, passengers who are starving on a lifeboat, can kill and eat a few of their fellows, provided the gains exceed the losses. But we often find such conclusions objectionable—and that intuition could influence the definition of problems for health reform.

3. Ethical Theory #2: Liberalism

A key objection to the utilitarian solution for the lifeboat problem is that some passengers are sacrificed for the sake of others. In the language of philosophy, the utilitarian logic treats some individuals as means and others as ends. For many, this approach violates fundamental values about individual dignity and equality. How can we legitimately sacrifice some for the sake of others, particularly without their consent? How can we decide that some lives are worth more than others without violating basic ideas of fairness? If these concerns constrain choices in the lifeboat example, then they may also have implications for health reform.

The most influential philosopher for the ideas of individual respect and autonomy is the eighteenth century German philosopher Immanuel Kant. According to Kant, all human beings have the capacity for *moral action* (Kant, 1788), the power to know what is morally correct, and to decide whether to follow the dictates of morality. Modern Kantians argue that since human beings have the capacity to develop and implement their own decisions about how to live—what philosophers call "life plans"—they have the right to do so. Because these rights derive from each person's status as a human being, they are seen as universal, and all political systems are obliged to honor them. This view directly opposes utilitarianism's willingness to treat some people as a means. Based on this approach, liberals create a set of rules that define how the state should operate and how policies should be determined (Rawls, 1971).

Kant argued that respect for all human beings required impartiality, as expressed in what he called the *categorical imperative*. This principle, which had several different formulations, was based on the general injunction to respect each individual as an end in themselves, not as someone to be used as a means to another's ends. What would such respect require of us?

The rights implied by the principle of mutual respect are interpreted by liberals in two different ways. *Libertarians* believe that only *negative rights* deserve protection (Nozick, 1974). These rights guarantee individual freedom, so that people can do what they want without state infringements on personal choice. Extended to the political realm, this formulation leads to fundamental political and civil rights, like freedom of speech, assembly, and political participation. Libertarians want only a minimal state to protect individual property rights and personal liberty. They typically oppose restrictions on drug use, limits on abortion, or even the licensing of physicians, since these are all actions that restrict individual freedom of choice.

In contrast, *egalitarian liberals* argue that the right to choice is meaningless without adequate resources. They argue that genuinely respecting others as moral actors requires us to provide others with the preconditions that make meaningful choice possible. Therefore, everyone has a *positive right* to the minimum level of services and resources needed to assure fair equality of opportunity. Someone who is starving, homeless, uneducated and ill does not have much opportunity for meaningful choice. The key

question for egalitarian liberals is what level of resources the state should provide to assure positive rights. For public health professionals, the question is how to interpret positive rights to health and health care. In the context of health reform, this question takes the form of asking whether there is a right to health care—or to health itself. Furthermore, should the violation of such rights be part of the problem definition that prompts health reform?

Another way to ask this question is, What level of health care are the poor entitled to? Egalitarian liberals disagree deeply on this issue. Some assert that the best way to respect everyone's moral capacity is to redistribute income and let individuals buy the health care (or health insurance) they want. In this view, health is no different from other goods and services that people purchase. For these liberals, there is no reason to worry specifically about the utilization of health care.

Others believe that society has a special obligation with regard to health. But here as well there is division. Some think that the key is providing a minimum level of *health care* for all, while others think that the critical issue is individuals' actual *health status*. In other words, if citizens have health rights, should we judge whether those rights are satisfied by the availability of clinics or by citizens' average life expectancy? The latter view holds that if society is to provide the prerequisites for moral action, access to care is not enough: what really matters is whether people live long enough, and at a high enough quality of life, to be able to choose among a range of life plans. This is the most common position among egalitarian liberal philosophers.

The disagreement over whether government should be responsible for health status or for access to health care has important implications for the relationship between the state and the individual. Much of ill health today depends on a person's own behavior. Smoking, diet, substance abuse, exercise and risk-taking of various kinds all have significant health consequences. To say that government is responsible for everyone's health status means that society has a responsibility to influence individual choices about such matters.

On the other hand, if society is only responsible for making services available, and not for whether people use them, then the individual has much greater responsibility for health consequences. Health then becomes a result of individual choice, not a prerequisite to it. This perspective is consistent with A.K. Sen's argument that society should be responsible for creating opportunities (which he calls "capabilities") that citizens can choose among, not for the choices that individuals make (Sen, 1999).

Even if we decide that individuals have a right to a minimum health status, we still must decide what that level is. Some people will be very ill, and it will be very costly to get them up to a reasonable health status goal. What are the limits of society's obligation to produce good health for the very ill? After all, society's resources are limited, and the money to produce such care comes from other citizens, in ways that effectively diminish their opportunities.

The minimum standard that a society picks for health status or health care will be influenced by a number of factors. These are likely to include its cultural and political balance, the level of economic development, and the availability and cost effectiveness of medical technology. Indeed, egalitarian liberals would argue that one of society's defining choices is how much it is prepared to spend on health care to preserve this basic aspect of individual opportunity.

3.1. Liberalism and Health Care Financing

In deciding how much to spend on health (or health care), a critical issue for liberals of all types is the legitimacy of taxation. For libertarians, negative rights include property rights, and taxation is therefore theft. In this view, not only do I have the right to be left alone, but my property is an extension of myself: the state does not have a claim on my property, beyond what is needed to provide minimal state services like defense and police. Taxation that seeks to redistribute resources in society is fundamentally not legitimate, because it treats one person (the taxpayer) as the means to another's end (the benefit recipient).

Egalitarian liberals, on the other hand, approve of taxation to achieve distributive justice, because they believe that much of the existing distribution of property is not legitimate: aristocrats acquired wealth through force and conquest, fortunes were made through monopoly or fraud, and some individuals today earn high incomes due to superior education or family contacts because they were lucky enough to be born into privileged social groups. Such gains were attained through the *social lottery* and, according to liberal egalitarians, these income differences are not deserved and hence are not legitimate. Therefore, property rights are not truly violated by taxing away such gains, because the gains themselves were not obtained in morally defensible ways.

In general, egalitarian liberals assert that an unequal distribution of income is morally defensible only if it is *effort sensitive* and *endowment insensitive*. Even differences in rewards due to differences in innate talent are illegitimate, since no one deserves their own biological advantages. The gains from such advantages are not the result of individual effort; income differences due to this *natural lottery*, like those due to the social lottery, can therefore be claimed by the state for redistributive purposes.

In health reform, egalitarian liberals assert that it is acceptable (indeed, desirable) to finance health care services with redistributive taxes. An egalitarian would believe that the more redistributive the financing, the more the health care system contributes to overall economic justice. Thus, income taxes are better than payroll taxes as a source of financing, especially since the latter often exclude investment income. To egalitarian liberals, income from capital is especially non-effort dependent and, hence, is especially illegitimate.

Egalitarian liberals view the U.S. health system as particularly objectionable. The U.S. finances most health care through private, premium-based insurance schemes in which costs are the same for all, regardless of income. As a result, consumption of health

care in the U.S., like the consumption of expensive consumer goods, such as sailboats, reflects an individual's ability to pay: rich people own large boats, middle-class people own modest boats, and poor people swim. The same is true when it comes to health care. Most of the uninsured in the U.S. are in low-income families: poor people cannot afford decent health insurance anymore than they can afford decent sailboats. This approach reflects a strong bias in favor of using market mechanisms and toward libertarian values.

3.2. Helping the Worst Off

How much redistribution should a society seek to provide? The most influential theory on this question was offered by philosopher John Rawls (1971), through a hypothetical thought experiment. Imagine, he said, that you had to choose the ethical principles for your world, and that you did not know who you might become. Imagine also that you were making this choice from a position of ignorance about your own particular tastes and values. In this hypothetical "original position," Rawls claimed, you would be risk averse, and would choose to make the worst off as well-off as possible, since you might become one of them. Rawls considered the ethical principles derived from this original position as the basis for distributive justice, or "justice as fairness." Since we would choose a world where the poor were as well-off as possible, we should design social policy to make the worst off as well-off as possible.

In this argument, Rawls did not treat health as special, but he did make a claim for extensive redistribution. It was not enough, in Rawls' view, to see that the poor met some minimum standard of well-being: as long as we can improve the condition of the worst off, we are obligated to continue redistribution.

This redistribution for the poor should be financed by raising taxes on the rich, until the tax burden slowed economic growth so much that the poor's bigger share of a smaller pie was less than the smaller share of the bigger pie they would receive if taxes were lower and growth higher. This issue is relevant today, especially in some Latin American countries and in Western Europe, where there is debate about the "Rawlsian Limit." Some politicians argue that the poor would be better off with lower pensions and less generous health insurance, because they would gain more than they would lose from the expanded economic growth such reforms would produce.

Any claim about the "unfairness" of the distribution of health or health care in a society is likely, at its root, to be based on egalitarian liberal ethics. For example, in a poor country, health spending in rural areas is well below that in urban areas, and life expectancy is shorter. Should this problem be a focus for health reform? The answer depends on the values of the person asking the question. For a utilitarian, whose goal is maximizing GDP, shorter life expectancy in rural areas might not be important—especially if there were a labor surplus in the countryside. To an egalitarian liberal, however, seeking to correct this inequality would probably be a priority for society.

Despite deep differences, there are some basic similarities between liberalism and utilitarianism. Both doctrines are universal: they seek to develop a single moral standard

for all human societies. In addition, they focus on the individual—on *individual* well-being and *individual* rights. Both perspectives have therefore been criticized for ignoring the social nature of human life (Sandel, 1982). Critics of liberalism argue that important community values are ignored by its individualistic vision, while utilitarianism is faulted for implying that you cannot favor your own family, friends or fellow citizens over strangers, if helping strangers would yield more utility (Williams, 1973). Similarly, utilitarian approaches to improve health status can conflict with a society's views about moral conduct. Consider, for example, the controversies in the U.S. over distributing clean needles to drug users or contraceptives to high school students (Moss, 2000). These criticisms lead to an ethical perspective that is not based on consequences or rights, but focuses instead on inculcating virtue and fostering community.

4. Theory #3: Communitarianism

Our third major ethical theory does not look at where people end up (utilitarianism) or at individual opportunities for choice (liberalism). Instead, it argues that what matters is the kind of society and kind of person state action seeks to create. This perspective focuses on the nature of the community and is called *communitarianism*. This theory argues that the character of a community depends on the character of the individuals who comprise it; the state should therefore ensure that individuals develop good character and help produce a good society.

Just as we examined two kinds of utilitarians (objective and subjective) and two kinds of liberals (libertarian and egalitarian), we distinguish two kinds of communitarians. The first are *universal communitarians*, who believe there is a single universal model for the good individual and the good society – their model. There are many examples, both secular and religious. The world's proselytizing, monotheistic religions (Islam and Christianity) are forms of universal communitarianism; so too is the position of ecological activists who want to transform man's relationship to nature; and some in the feminist movement (who want to create a society with a different relationship between men and women) are also universal communitarians. The same can be said of revolutionary Marxists.

An important non-Western form of communitarianism is offered by the Asian philosophy of Confucianism. This perspective, articulated through the writings of Confucius (551-479 B.C.) and his followers, prescribes how people should behave, both as individuals and as collectivities, through the "five relationships." Human interaction begins in the family and extends out to the state, providing a distinct Confucian continuity between these two realms of collectivity. The *Four Books of Confucianism* provides a universal view of humanity, based on internal reflection on one's duties and rights, and on understanding one's human interactions within specific collectivities (Bloom, 1996).

The Confucian arguments about duty and responsibility, and about fulfilling social roles, thus resemble the arguments offered in the Greek philosopher Plato's famous work, *The*

Republic, but with important differences. While Plato viewed the family as a particularistic distraction from the social obligations of citizenship, Confucius considered the family as the basic model of human interaction, for ruler and subject, for public and private matters (Bloom, 1996, p. 134). Confucian ethics continues to influence individual and social values throughout Asia. The willingness of the government of Singapore to use the state to preserve order and instill individual virtue is an expression of this perspective today.

Universal communitarianism thus allows for much more substantive variation than either liberalism or utilitarianism. Unlike those nineteenth century ideas, the focus of communitarianism is not on isolated individuals, but on the nature and structure of society as a whole. And the substantive ethical ideas of what constitutes the good society can vary widely for communitarians.

The second kind of communitarianism responds to the wide variation in visions of the common good, by recognizing that each community should decide its own norms and organization. This leads to *relativist communitarianism*. The contrast with liberalism is instructive. For liberals, all individuals should be free to decide for themselves what makes a good life and to pursue their distinct visions. Relativist communitarians, on the other hand, do not believe in leaving individuals free to pursue their own visions of the good. Instead, they believe that the community should seek to instill particular notions of virtue, character, and goodness into its citizens, and the community (or the state) can legitimately coerce individual members to follow the traditional values.

Certain practices, like female genital surgery, pose the question of community norms versus universal norms in a sharp manner. One position is to oppose such practices on the objective utilitarianism grounds that they can injure the health of women. On liberal grounds, the practice is objectionable because it can limit long-term life opportunities. But relativist communitarians would support the practice, arguing that such values should be respected by outsiders, because it has deep meaning for people who belong to the culture.

Someone who believes that each community should define its own norms must decide on the boundaries of community and on who speaks for the community. Many ethnic and religious minorities have asserted the right to depart from the norms of the larger community and the right to create their own communities. How are resulting disagreements among communities to be decided? Can Catholic majorities in Latin American countries impose their views about divorce and abortion on all citizens, on the grounds that this is what the community believes? What rights do religious minorities have in Israel, India or Iran, if their practices offend the majority? Suppose people of the same sex want to marry or adopt children? Until twenty years ago, women in Switzerland could not vote, and the Swiss defended this limitation as their traditional practice. Who decides whether a community is legitimate, and what are the acceptable limits of coercion for a community to use in promoting its values and assuring compliance by dissenting members? These are critical questions for a relativist communitarian perspective.

Communitarian arguments arise in health reform in many ways. For some in public health, living a healthy lifestyle is a matter of virtue, not just a matter of cost effectiveness. For example, some public health advocates view tobacco or intravenous drug use as ethically objectionable forms of self-destructive behavior, not just as unhealthy or unwise. Advocates of these views refuse to accept others who prefer the pleasures of tobacco or reject the joys of exercising. These choices would be acceptable to a subjective utilitarian (the smoker is seeking happiness) or to a liberal (the smoker has rights), but are often opposed by a public health communitarian (as not following the right kind of life).

Communitarians may also have strong views about particular health care services. In the U.S., many religious conservatives oppose the availability of abortion or family planning services on communitarian grounds. In some countries (e.g., Japan), the practice of defining death by the end of heart action limits the number of organs available for transplant. Should public health professionals, arguing on objective utilitarian grounds, seek to change such community norms? Some countries (e.g., Austria) presume that those who die are organ donors unless the family actively opposes it, a practice that has obvious health-status gains. Should other countries follow suit, even if doing so runs counter to local tradition? Questions around traditional medical practices raise similar issues. If a sincere believer visits a religious healer who seeks to expel evil spirits to treat a medical problem, should the healer be required to follow guidelines for medical practitioners, as stipulated by law?

In offering these examples, we seek to illustrate how using community norms to define the agenda for health reform often confronts conflicts among different values and ethical perspectives. For example, if the rich have much better health care than the poor, is this a problem for health reform? What if the country's constitution guarantees a right to health, but the majority of people in the country believe in values of independence and free enterprise, which could justify inequality of health?

This chapter has emphasized the normative dimensions of health reform, asking how a society *should* decide what to accomplish in health reform. From that perspective, reformers will have to decide on the appropriate role for communitarian visions, and the relationship to utility-based and rights-based values.

5. The Problem of Justification

Confronted with these three basic ethical positions, a reformer might well ask which one is "correct"? How can any particular ethical claim be defended? What arguments are available to select one ethical view over another? Such questions about justification fall into the realm of meta-ethics, or questions about the nature of ethics itself.

Various approaches exist to the question of justifying an ethical position. Religious faith is one. Another is the view that human beings have a special faculty for

perceiving morality, so that moral truth is revealed to us by our emotions or intuition. A third view holds that logic or reason can dictate the content of morality (Harman, 1977). The most widely accepted modern argument is "moral realism" (Putnam, 1990): the position that the content of morality can be learned from our experience, if properly understood and processed. All three of our ethical perspectives above—utilitarianism, liberalism, and communitarianism—belong to this view, although they offer different answers on the substance of morality and the method of justification. Moral realists believe that moral truth exists, that it can be justified, and that it can be learned through understanding human nature, by analyzing human needs, or from the requirements of social life.

Other contemporary thinkers reject all these lines of analysis. The *postmodern* school argues that ethics is not justifiable in any foundational way (Lyotard, 1984). Instead, postmodernists argue, ethics is created, like art or poetry. Criteria for judging moral arguments are based on rules internal to the enterprise, like the stylistic norms that govern an artistic tradition, and they cannot be derived from more fundamental principles—they do not and cannot have a deeper justification. The words that human beings use to describe concepts like "justice," "well-being" or "tradition" are just that—words—symbols that express ideas invented by people, just as Gothic architecture and country music were invented by people.

A postmodern health reformer would be liberated from the problems of justification, but would still need to grapple with making moral decisions. If there is no fundamental justification, then are all moral views and all courses of action equally compelling? Richard Rorty, a prominent American postmodernist, argues that moral judgments are possible, even though they cannot be justified in a foundational manner. We have no choice, he says, but to act on our own view of the good, and to seek to persuade others to accept our perspective. For example, Rorty is prepared to advance the cause of respect for human rights, even though he cannot prove that this moral position is "correct" by reference to a higher law or basic principle. Instead, Rorty urges a modern version of the doctrine of pragmatism, argued by John Dewey (1929). Rorty proposes the acceptance of moral views that "work" to make the world a better place, as best we know how. Rorty's personal choice is a form of egalitarian liberalism, and he seeks broader acceptance of this moral perspective via "poetic" or "prophetic" means, since he recognizes that rational argument alone will not suffice.

6. Developing An Ethical Position

We have presented these ethical theories as if they were mutually exclusive—as many philosophers have done before us. But such a purist perspective may not be appropriate for our purposes. Indeed, the various theories can be interpreted as embodying complementary rather than competitive insights. For purposes of health reform, it may make more sense to seek a "mixed" position, drawing on the intuitions behind several of the ethical perspectives we have reviewed.

How can a serious reformer develop an ethical position that is appropriate for public policy yet compatible with personal values? One approach is to work both "forward" and "backwards"; that is, one can start with an appealing theory and explore its implications for practice. One can also do the reverse: begin with intuitions about desirable practice and work back to the theoretical stance that would support the policy position. In this way, both specific intuitions and more general theoretical ideas can be modified through an iterative process, in pursuit of the consistency that philosophers call "reflective equilibrium."

This iterative process will reveal the limits of any one theoretical position. For example, an individual committed to an objective utilitarian view of health maximization will likely find that she would like to have limits to the level of coercion employed to reach that goal. Notions of liberal respect and of the "negative rights" to autonomy can help clarify the theoretical explanation of these limits to coercion for specific policy impulses. This kind of mixed position might be called "rights-constrained utilitarianism."

Mixed views can also be reached via arguments that show how, in certain specific situations, one theoretical stance leads to the acceptance of an apparently different ethical principle. For example, subjective utilitarians would not want to overturn local cultures and customs (a relativist communitarian concern), if preserving these local values promoted happiness among a community's members. On the other hand, a relative communitarian living in an egalitarian liberal society would be obliged to accept these liberal norms as the basis for day-to-day substantive policymaking.

Our attitude toward mixed ethical positions is rooted in our view of ethical theory more generally. We are acutely aware of the simplified terms in which ethical issues are normally discussed in public policy debates. Even the words used to analyze health policy—words like "rights" or "well-being" or "community" or "health" itself—map imperfectly to the rich and complex texture of actual problems. Like pictures of the same landscape painted by painters with sharply different styles, different ethical theories capture distinct perspectives on any one scene, calling our attention to different features and patterns. Each rendition can contribute to a fuller appreciation of the overall terrain, depending in part on the viewer's capacity to assimilate each perspective.

We do believe, however, that health reformers should seek consistency and coherence in their construction of ethical positions for changing the health system. That is, we should strive to do more than just picking and choosing from various theories as whim moves us, in what could be styled "buffet ethics." If that is all we do, then ethical theory becomes window dressing, a fancy rationale for a previously chosen position, without any analytical force.

The construction of a mixed ethical view, therefore, should provide a coherent and persuasive rationale. Take our earlier example of someone who believes that the pursuit of utility is to be limited by certain rights. This person should be able to explain why the boundary between these two principles is drawn where it is, and defend to other people why they should accept this particular mix of utilitarian perspective and rights.

Coherence and explicitness are important for several reasons. First, they make one's views more explainable, which is critical for public debate about the ethical dimensions of health-reform proposals. Second, coherence and explicitness create the potential for agreement and disagreement, allowing others to modify or defend the ethical position more effectively than an ad hoc "mush" of positions. Finally, the development of an ethical stance for health reform is an ongoing process and not a "once and for all" activity. The complexity of health system problems and the limits of ethical analysis categories help ensure that the process will continue. As reformers move around the "cycle" of health sector reform, discussed in Chapter 1, they should seek to learn and grow ethically as well as scientifically and politically.

7. Summary

Each ethical theory we have considered in this chapter has its characteristic questions which reflect unresolved issues:

- For utilitarians, how should well-being be measured?
- For liberals, which rights do people have?
- For communitarians, what are the boundaries and values of the good community?

Each ethical theory provides important insights into the human condition and can guide the normative decisions that must be taken in health reform. Utilitarians focus on consequences—where people end up; liberals focus on rights and opportunities – where people start; and communitarians focus on the kinds of individuals and communities that we seek to create.

Everyone involved in health reform must decide for themselves what their personal values are, and how far they are willing to use coercion to advance those values. This process raises difficult questions when political institutions and power structures produce answers that an individual disagrees with: How far can one go to discuss, protest, or undermine such decisions? The answer depends importantly on the quality of the process that produced the decision. The more accountable, open and democratic the process, the more worthy the answer is of respect, even from dissenters.

References

Bentham, Jeremy. 1789. *An Introduction to the Principles of Morals and Legislation*. Edited by J.H. Burns and H.L.A. Hart. New York, NY: Oxford University Press, 1996.

Bloom, Irene. 1996. Confucian perspectives on the individual and the collectivity. In: Irene Bloom, J. Paul Martin, and Wayne L. Proudfoot, eds., *Religious Diversity and Human Rights*, New York, NY: Columbia University Press, pp. 114-151.

Bodenheimer, Thomas. 1997. The Oregon Health Plan – lessons for the nation. *New England Journal of Medicine* 337: 651-5, 720-3.

Dewey, John. 1929. *The Quest for Certainty: A Study of the Relation of Knowledge and Action*. New York, NY: Putnam, 1960.

Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. New York, NY: Oxford University Press.

Kant, Immanuel. 1788. *Critique of Practical Reason*, English trans. by L.W. Beck. New York, NY: Liberal Arts Press, 1956.

Lyotard Jean Francois. 1984. *The Postmodern Condition: A Report on Knowledge*. Trans. By Geoff Bennington and Brian Massumi. Minneapolis, MN: University of Minneapolis Press.

Moss, A.R. 2000. Epidemiology and the politics of needle exchange. *American Journal of Public Health* 90:1385-1387.

Murray, Christopher J.L. 1994. Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization* 72(3):495-509.

Nozick, Robert. 1974. *Anarchy, State and Utopia*. New York, NY: Basic Books.

Putnam, Hilary. 1990. *Realism with a Human Face*. Cambridge, MA: Harvard University Press.

Rawls, John. 1971. A Theory of Justice. Cambridge, MA: Harvard University Press.

Rorty, Richard. 1989. *Contingency, Irony, and Social Solidarity*. New York, NY: Cambridge University Press.

Sandel, Michael J. 1982. *Liberalism and the Limits of Justice*. New York, NY: Cambridge University Press.

Sen, Amartya 1999. *Development as Freedom*. New York, NY: Knopf.

Weinstein, Milton C., and William B. Stason. 1977. Foundations of cost-effectiveness analysis for health and medical practices. *New England Journal of Medicine* 296:716-21.

Wenz, Peter S. 1986. CBA, utilitarianism and reliance upon intuitions. In: George J. Agich and Charles C. Beyley, eds. *The Price of Health*. Boston, MA: D. Reichel Publishing Co.

Williams, Bernard. 1973. A critique of utilitarianism. In: J.J.C. Smart and Bernard Williams, eds. *Utilitarianism For and Against*. New York, NY: Cambridge University Press.

World Bank. 1993. *World Development Report: Investing in Health*. New York, NY: Oxford University Press.

World Health Organization. 2000. *World Health Report*. Geneva: World Health Organization.

Chapter 4
Criteria For Health Systems Evaluation

1. Introduction

How can we go from the philosophical discussion of the previous chapter to practical judgements about the performance of health care systems? Those who subscribe to different ethical theories will not necessarily endorse the same criteria for evaluating that performance – nor place the same importance on various performance measures. Does that mean that all we can say about such judgements is, "It depends?"

While all performance judgements are value-relative, we believe that health sector reformers can make defensible decisions about criteria to use to both define problems and evaluate solutions. In this chapter, we propose just such a set of criteria, and discuss their meaning, limits, and utilization. We review their connection to philosophical theory and explore both what is included and what is left out of our formulation.

Our basic argument is quite simple. The health sector is a means to an end, not an end in itself. Therefore such a system should be judged by its impact. We will identify three variables – which we call *core performance criteria* – by which to evaluate the health sector. These are: 1) the health status of the population, 2) the satisfaction citizens derive from the system and 3) the degree to which citizens are protected from the financial risks associated with ill health. We will also argue that cost considerations in various ways are always a part of the problem motivating health sector reform. (Readers familiar with the WHO *World Health Report 2000* will be aware that that document too proposes three quite similar criteria. However, as will become clear as our exposition proceeds, we have important disagreements with that formulation with regard to overall approach and numerous specifics instances)

We will proceed as follows. First we discuss our basis for choosing criteria, and the relationship between health sector reform and social and economic policy in general. Next, we will explain and defend our core performance criteria, and explore the role of cost in the process of problem definition. Then we describe certain *intermediate criteria* – variables of interest because of their impact on our core criteria. We go on to discuss how the process of problem definition has to respond to the cultural context which both shapes concerns and limits possible responses. The last section briefly synthesizes what has come before, particularly with regard to being strategic in selecting performance problems for priority attention. Throughout the chapter, we will assume that a particular person (such as the reader) is thinking about health sector reform, and our advice is directed at the decisions and choices that such an individual has to make.

Our proposal is, of course, influenced by our personal judgments, experience, and values. Picking performance criteria is not a mechanical process. We have tried to be open and self-reflective in presenting the performance criteria in this chapter. Since we cannot escape the force of our own ethical concerns, we try to acknowledge these as we

proceed, so that those whose judgments differ from our own can still use our framework to pose evaluative questions clearly and explicitly.

2. Using and Choosing Criteria

To sensibly pick performance criteria, an analyst has to know the purpose for which they will be used. For example, the WHO's recent *World Health Report* ranked all countries' health care systems. This required WHO to construct numerical measures that could then be combined via a simple formula. Our task is different. We want to use performance criteria to identify performance problems for priority attention in carrying out health sector reform. As we will see, this makes numerical scores less useful, and detailed descriptive data more important, for our purposes.

To reiterate a point made in Chapter 2, in our view, these measures should reflect the *results*, or *consequences* or *outcomes* of the health sector. The "policy cycle" should begin with identifying how the performance of the health sector is inadequate. This insistence, on starting with outcomes, is one of the distinctive features of our approach. A reformer then can go on a "diagnostic journey" (see Chapter 6) to figure out the causes of that inadequate performance. Only then is it possible to devise sensible reform strategies to improve that performance. (The same set of criteria can also be used to evaluate policies, once they are implemented.)

In using the term *problem* to mean *performance* problem we are calling for a somewhat specialized usage. Physicians, in contrast, often do use the word *problem* to mean *cause* – when they tell a patient, "Unfortunately your problem is that you have cancer." Many health sector reformers also do not think in the way we recommend. Instead, when asked, "What is the problem is your country?" they immediately answer in terms of some feature of the system: "We don't have enough primary care" or "We need to introduce social insurance." We, in contrast, contend that such statements are actually about potential *causes* of (or potential remedies for) performance problems – not definitions of the underlying performance problem. How can one know if a nation's physician workforce is inappropriate or its financing not properly organized unless these features of the system produce unacceptable results?

This leads us to our next question – namely how do we judge various performance measures once we have data on them? In particular, for a given population, what parameters of the distribution of experience – if any – are we interested in? For example, some have argued that both a measure of average performance, and of the equity of the distribution of performance, are relevant for each criteria.

Again we propose to make decisions based on our purpose – mainly guiding real health sector reform. To generate comparative rankings of countries one or two simple summary statistics are all that is required. But a real reformer will derive little help from such numbers – except perhaps for political or public relations purposes. Yes, a nation's average performance on a measure like life expectancy can help focus attention on the

health sector through the process of international benchmarking – or by comparing the nation's current situation with its own past performance. But for these purposes, comparisons among various alternative statistical measures (mean, median, mode) involve a level of detail that is likely to be beyond the sophistication of all but a few technocrats. The same is even more true of measures of variation in outcome within the population. While the "average" has some intuitive meaning to many – few will have any intuitive associations that allows them to relate to measures of variation like the variance, the inter-quartile range or the Gini coefficient.

We do believe that in judging a health system, the *equity* of outcomes should matter, along with *average* performance. Thus we urge reformers to think in terms of a three-by-two matrix: three performance criteria and two ethical tests to apply to each. But when it does come to matters of equity, we are painfully aware that different nations not only *value* equity differently, but that even more basically they *define* equity differently. For example, some see any difference in service use between rich and poor as a problem (e.g. Denmark) while others focus on bringing up the bottom while allowing the top to get more if they both want to and can afford it (Germany, Australia). Hence the norms countries use to evaluate performance inside each cell in the matrix have to come from them. Nor is it impossible to imagine that a particular country might have different distributional values for different performance criteria—for example, being more concerned to produce equity in health outcomes than in satisfaction.

Moreover from a reformer's point of view, it is unlikely that any single statistical measure of equity will be very informative. Instead, as we stress in what follows, relatively detailed information about the distribution of outcomes – across regional, income or ethnic groups – is generally most relevant. For in practice, efforts to reduce inequality usually turn on bringing up the bottom – not bringing down the top. And to do that, those who experience relatively poor health, satisfaction or risk protection need to be identified in policy-relevant terms.

Indeed, as we will discuss further in the chapter below, and in Chapter 6, data on performance generally need to be disaggregated not only by population group but also by the various causes that contribute to overall results. If life expectancy is low and/or declining – what are the causes of death that are contributing to that result? If citizens are unhappy, what aspects of the services they receive do they find objectionable? If some are impoverished by health care costs – what are the circumstances under which that occurs? In a sense then, we are urging a return to the research program followed by Murray, et al. when they compiled detailed data on disease burden by disease – with the important caveat that even that effort imposed needlessly uniform "weights" to add up and combine different kinds of health consequences. And we would also emphasize the need for disaggregated outcomes data in order to both make equity judgments and, more importantly, to guide equity-promoting action.

To choose a particular set of performance criteria, then, is to choose a particular focus for public debate about health sector reform. What basis then should we use for choosing such performance measures? Our first basis for choosing such measures is

political relevance. We seek performance criteria that will be helpful to practitioners, as they struggle with the politics of health reform in their country. We have a firm belief in both the relevance and legitimacy of politics in the reform process. Therefore, the performance criteria we select should reflect and embody the major political and social concerns that have driven (and are driving) health sector reform in various nations around the world. Reformers have to be able to recognize and express their concerns within and through the framework we offer. The process of deciding how to use our criteria should allow practitioners to locate themselves in the current debate. In working with many governments around the world, we have encountered certain common issues or concerns. And our criteria should capture these concerns.

Our second basis is that the core performance criteria should be relevant to the main lines of philosophical concern about the health system, as explained in the previous chapter. To be truly useful, performance measures should reflect those various important ethical positions. Utilitarians, liberals and communitarians should all find their concerns represented. Otherwise they will not see the framework as useful to them. Taken together, the set of measures should capture the critical ethical issues at stake in the reform process. Different views on those issues should be expressible in terms of disagreements about the relative importance of, or appropriate definition of, the criteria in question.

Our third basis for choosing criteria has to do with causal dependence. We want performance measures that are significantly affected by health policy choices. For example, while the health sector does influence citizens overall level of happiness, such general well being depends much more heavily on a wide variety of events and circumstances outside health care. On the other hand, satisfaction with the performance of the health sector is largely (although of course not entirely) dependent on what happens inside the sector. Hence the second, narrower aspect of satisfaction meets the test of causal dependence while overall levels of happiness or well being do not.

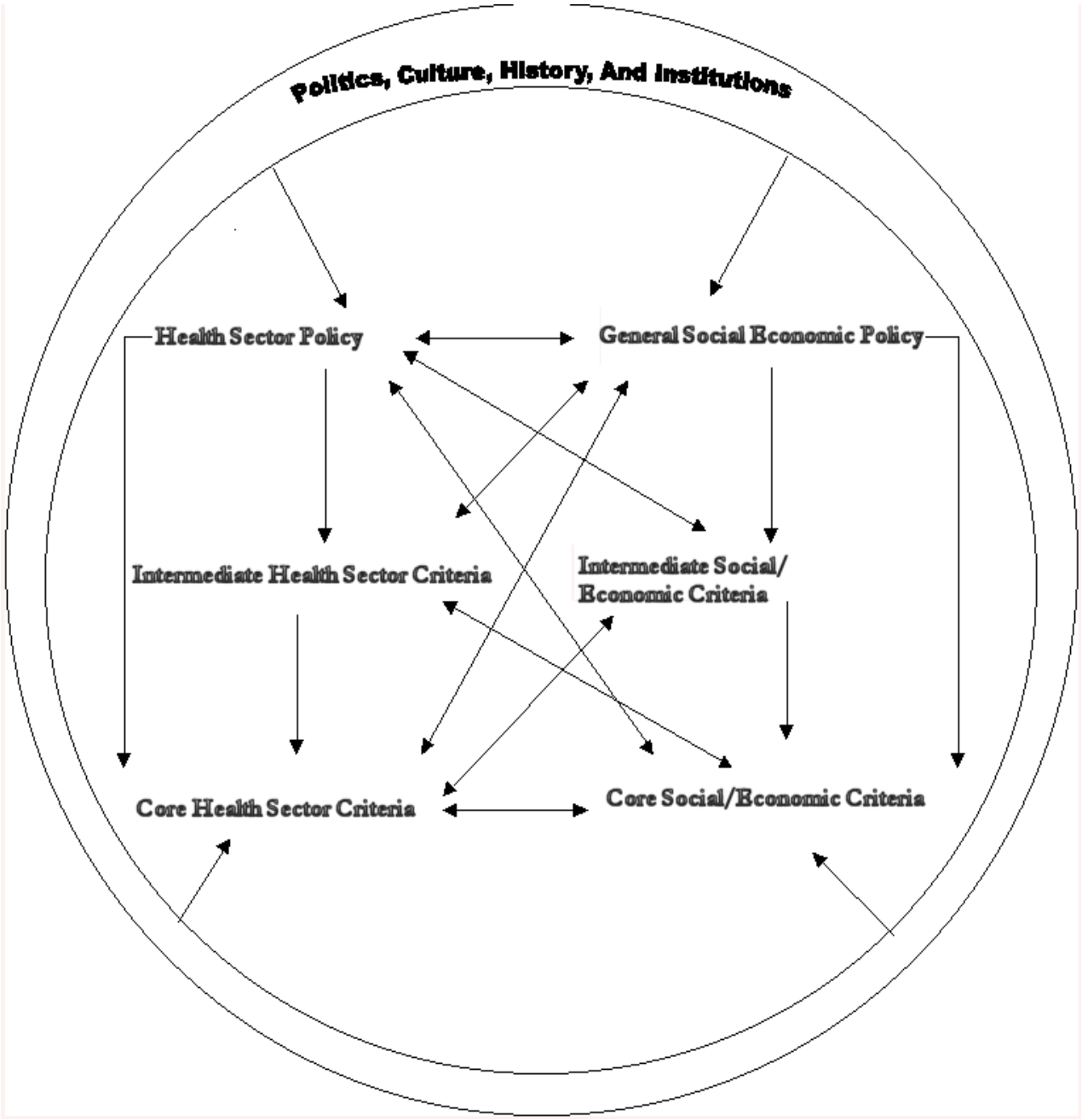
As we will see, reducing any set of performance measures to usable form requires many additional decisions about definition, measurement, etc. These decisions will also necessarily reflect an analyst's particular values and priorities. So, once agreement has been reached on certain core criteria, the detailed specification of these indexes is likely to reveal and/or provoke further discussion that will raise serious ethical choices – choices that not everyone will agree upon.

3. Health Policy and General Social and Economic Policy

Health policy choices can have a direct effect on health sector performance, and they also can act indirectly – influencing certain "intermediate" variables that in turn have an impact on that performance. But these two pathways are not the whole story. The government does many things in the domestic policy arena outside of health. For simplicity's sake let's call these activities the "Social and Economic Policy System." And that system, and the health sector, interact in a variety of ways.

Figure 4.1 depicts both the health sector and the general social and economic system, as well as the relationships between the two. The general system has the same schematic form as (but different content from) the health system. The resulting picture is rather complicated. All the variables in each system potentially affect all the variables in the other, as Figure 4.1 indicates. In addition, all these variables are themselves influenced by a variety of more general factors – which we have simply labeled "Politics, Culture, History, and Institutions" in Figure 4.1.

Figure 4.1.



These interactions flow in both directions. For example, tobacco taxes imposed to lower smoking could adversely affect economic growth in tobacco growing regions. Conversely, general social and economic policy developments (for example, a government's commitment to regionalization) can lead to changes in the health sector. Thus one of the bases we have used for choosing intermediate criteria is to ask which aspects of health sector performance have an important effect on non-health outcomes. Simply put, health sector policy cannot always be evaluated in isolation.

Consideration of the political and social context leads us to expand our evaluative focus in another way. Both intermediate and core criteria describe circumstances inside the health sector itself. To take account of the reciprocal connections between the health sector and the rest of society depicted in Figure 4.1, two major contextual factors need to inform health sector reform efforts. One of these is the cost of the system, since that affects the burden the health sector imposes on society and how much money the society has to pursue other goals. The other involves the relationship of reform to community norms and cultural practices in the society in question. We discuss both of these relationships below.

The full picture can be summarized as follows. The three core health sector performance criteria we will propose are: 1) health status, 2) citizen satisfaction, and 3) financial risk protection. The four intermediate criteria we consider are: 1) access, 2) quality, 3) efficiency and 4) financial burden. In addition, we explore two aspects of the broader reform context: cost and community values. We list all of these factors here so that partisans of one or another of these measures will not think we have neglected or omitted them. Instead, we will examine each in turn – and explain why we have located it in a particular place in our overall framework.

4. Core Health Sector Performance Criteria

4.1. Health Status

The health status of the population clearly deserves recognition as a core performance criterion. It meets the test of political relevance since it is the subject of much public debate as well as policy making. It also meets the test of philosophical relevance. It embodies the central concern of objective utilitarians, and is a critical aspect of individual opportunity for many egalitarian liberals. It also meets the test of causal dependence. While other factors (e.g., economic status, and education) do affect health status, the operation of the health sector surely has a major impact on health itself.

As noted in Table 4.1, to construct an operational measure of health status, a policymaker has to make many specific decisions. And different nations choose different answers to these questions. For example, not all countries will count years of life lost at different ages in the same way, nor place the same relative value on current versus future benefits (i.e., use the same discount rate). We believe, however, that all reformers need to begin by asking this question: what will the impact of proposed reforms be on the

health status of the population? This question is important regardless of how they choose to measure that impact.

Table 4.1:
Decisions Required to Construct A Health Status Index

▪ The relative effect of different kinds of disability and disease. For example, in measuring disability, does only functional impairment matter, or are "silent" physiological conditions also relevant?
▪ The relative value of years of life lost at different ages.
▪ Do the economically productive or socially valuable matter more than the unemployed or retired?
▪ What impact, if any, should a person's non-health status have on the value of their life? Are some illnesses less serious when they happen to the wealthy because the wealthy can buy compensating services?
▪ Should attitudes of the public toward different diseases/causes of death matter? For example, should the fact that in the U.S., the public is especially eager to avoid cancer deaths, affect our valuations as to the gains from death prevention?
▪ How do we combine death and disability? Are the non-disabled's lives more valuable than those of the disabled?
▪ Are future gains to be "discounted" (i.e., values less than present gains), and if so, at what rate?
▪ What about uncertainty? Do we just make our best guess of average effects—and ignore our differences in the degree of confidence that these will occur—or should we not be so risk neutral?

A second way to think about health status performance problems is to identify those diseases that are causing the greatest harm. If a country is losing QALYs to alcoholism, or to a rise in T.B., or to neo-natal tetanus, then those conditions might well be a focus of reform efforts. In countries where measures like QALYs or DALYs are not available, targeting those conditions where incidence or prevalence are high relative to other comparable countries (or relative to a nation's own past performance) can be a straightforward way of using health status data to guide reform efforts. (The question of whether or not a country can do anything about such problems is also relevant to its priority setting – as we discuss below.)

As we argued above, statistics describing the population's overall health status (like average life expectancy) offer only limited guidance for health sector reform. Information about the *distribution* of health status is often more helpful. For example, are certain regions, ethnic groups or demographic groups doing worse on life expectancy than others (e.g., tribal peoples in India or the island provinces in Malaysia)? In effect, this process uses the country's own average performance as a *benchmark* for identifying performance problems that need attention. (We elaborate on the various forms of benchmarking further below when we discuss priority setting in general.)

Nations do have to decide how important equity is to them. That is, how important is it to raise up those with the worst health status, compared to improving the experience of citizens more toward the middle of the distribution? Indeed, one of the virtues of identifying health status as a core criterion is that it focuses attention on the question of exactly what aspects of that distribution are ethically important. And answers

to that question will be (and should be) shaped by each nation's own values and political process.

As egalitarian liberals we do think there are strong ethical arguments for paying special attention to the health status of the worst off. And sometimes (although not always) extending basic services to the under-served is also an efficient way to improve overall (average) health status – because primary care is often very cost-effective. Indeed there is some evidence that countries with very unequal levels of service (e.g. Turkey) have lower health status than other countries with a similar level of health spending. This is because – at the margin – the gains from added spending to those who have the most are less than the gains that would be realized by spending the same amount on those who have the least. On the other hand, sometimes those with the poorest health status are especially difficult to serve – because they live in remote areas or struggle with a variety of social and economic handicaps. Moreover not every country's political leadership shares our particular values. And as realists, we recognize that various considerations – from pleasing national elites to increasing a country's international status can produce political pressures to spend more on tertiary institutions. And minorities and the poor may lack the political power needed to ensure a response to their needs.

Our point, however, is that our views about health equity are not what matters. Instead reformers need to clarify *their commitments* on these issues, in order to know what problems to focus on and what reforms to advocate. Is it acceptable for the rich to buy better care for themselves – as long as the poor have access to some minimum? Does "fairness" require the state to pay for the poor to have access to the same expensive, life-saving technology that the rich do – even if that technology is not a cost-effective way to produce health status gains? And just how important is it to help the worst off, even if doing so is expensive? In poor countries, where resources are very limited, these can be especially difficult questions. But reformers must seek answers to them to evaluate the equity of the distribution of health status in their country.

4.2. Citizen Satisfaction

Our second core performance criterion is the degree to which citizens are satisfied with the services provided by the health sector. Philosophically, such a goal is in keeping with the subjective utilitarian view favored by economists. Politically, the system's inability to provide what citizens want is often a significant driver of reform. Moreover this goal allows us to capture various features of the health system, apart from its impact on health status. For example, how accessible and service-oriented is the care process? By using satisfaction as a core criterion we take account of how citizens themselves evaluate and respond to their care.

Again our approach departs from the recent WHO formulation, which considers only "legitimate" satisfactions.[1] Our reasons for rejecting that approach are both philosophical and practical. Philosophically, satisfaction is a subjective utilitarian concern. Within that framework, there is no basis for assessing someone else's satisfactions according to our

[1]

view of their "legitimacy."[2] Moreover, satisfaction with the healthcare system depends on various features of that system that interact. We cannot divide up "satisfaction" and then attribute parts of that total to various distinct bits and pieces of the consumer's experience.[3]

We also seek to avoid a process that would portray the decision to discount certain citizen attitudes as a mere technical decision about measurement. Our concern with political accountability implies a need for transparency about such judgements. If reformers in a particular country decide *not* to respond to some citizen desires, let them say so explicitly and argue for their choice in an open manner.

Of course there can be tradeoffs between increasing satisfaction and other goals. For example, patients might get satisfaction from inappropriate care – like unneeded injections. Responding to such desires can lead to lower health status and /or increased costs. Also those with poor health habits might argue that no effort should be made to change their behavior since those habits increase their own satisfaction. The decision to suppress or avoid such conflicts (by editing out what some reformer believes are 'illegitimate' satisfactions before making tradeoff decisions) seems to us an awkward and obscurantist approach.

Measuring satisfaction – however defined – is not easy. The economist's solution is to measure individuals' *willingness to pay* for various kinds of benefits. For example to put a value on human life, individuals have been asked hypothetical questions about their willingness to take specified risks in return for various payments (so-called "contingent valuation" studies).[4] These studies have revealed systematic inconsistencies between the behavior that decision theory says is rational and typical patterns of choice. Another example of such difficulties, noted in Chapter 3, is that ill persons may rate their quality of life as higher than when they were well.

An alternative approach to determining satisfaction is to use some sort of satisfaction survey – asking what people do and don't like about the health care system. Such surveys do not yield monetary values that are directly comparable to cost estimates. But they can provide valuable guidance to reformers, particularly if they elicit information about satisfaction with specific services or particular features of the system. Such studies do face certain technical problems (like the tendency of respondents – especially low status respondents – to answer "yes" to questions). But extensive work has been done in both Europe and the U.S. to develop and validate reliable survey instruments.[5]

Judging satisfaction, like judging health status, does require a country to consider the equity of the distribution of consequences – as well as the overall average result. Are gains in satisfaction to the happy and the unhappy equally important? Is the

2

3

4

5

dissatisfaction of those with good health status more or less significant than the negative reactions of those who bear a higher burden of disease? Subjective utilitarians, rooted in economics, would argue that total satisfaction – and not its distribution – matters.[6] For egalitarian liberals, a situation in which some are much more satisfied with how they are treated than others violates notions of equal respect. And as in the case of health status, as a practical matter, reformers concerned about the distribution of satisfaction need to identify what groups in the population are particularly unsatisfied.

In summary, we believe that citizen satisfaction with the health system should be a core performance criterion, despite methodological and philosophical issues. Citizen satisfaction is heavily influenced by the system, widely discussed, philosophically grounded, and politically relevant. The relevant metric, we have argued, should be all reactions of citizens, regardless of whether experts like or approve of these reactions or not. When this goal conflicts with other goals, or where citizens' reactions are ethically problematic, those issues will have to be confronted and addressed by health reformers. But since the problem of health sector reform is seldom neat or easy, this result should not be surprising.

4.3. Financial Risk Protection

Financial risk protection is a major goal of health sector policymaking, and a major focus of the politics of health reform. It is also greatly influenced by how the sector is financed. In addition, preventing financial impoverishment – and its associated loss of opportunity – is philosophically important to egalitarian liberals. For them insuring financial opportunity is as important as the prevention of early disability or death, in achieving their vision of positive rights.[7] In short, the reasons for including financial risk protection as a core criterion are compelling.

Providing financial risk protection, however, does not involve protecting the population against all the costs of health care. In fact that cannot be done. Foreign aid aside, all health care costs are ultimately paid for by members of the society – directly or indirectly. It is simply not possible to protect those in the middle of a country's income distribution against the costs of routine medical care. If they don't pay for those costs directly, they will do so indirectly via various taxes. What is relevant for this performance criterion is helping people avoid the risk of large unpredictable losses from an unexpected illness – that is, to provide a risk-spreading or insurance function. That is what we mean by financial risk protection.[8]

Measuring the extent of risk protection is complicated by the fact that the impact of a financial risk depends on both the size of the risk and on the economic status (income and assets) of the person incurring the risk. The closer someone is to the poverty level, the smaller the expense that will put them below that line. Hence the more they need financial risk protection.

6

7

8

An additional complication is that financial risk can also affect health status. Those who are seriously ill may not receive adequate health care if the financial burden of doing so is too great. Hence just because people do not incur costs does not mean they are not injured by a lack of risk protection. That injury may show up as decreased health status and not in the financial arena.

How then can the extent of financial risk protection be described? The simplest data – e.g., the extent of insurance coverage – are only partially informative. Some nations don't provide health insurance. They rely instead on providing free (or nearly free) government health care. Those systems provide some risk protection – depending on the accessibility and quality of the services. But when consumers don't fully trust public care, they may feel compelled to use personal funds to purchase private care when illness occurs.[9]

Insurance coverage, moreover, is not a "yes-or-no" variable. Many people with social insurance in developing countries are actually underinsured, and hence face significant financial risks if serious illness occurs. It is also possible for some citizens to have "too much" insurance. For a subjective utilitarian, there comes a point where individuals would prefer to spend more on other goods and services and forgo added risk protection. In a free market, not everyone purchases maximum insurance against all possible contingencies. Objective utilitarians also can find examples of over-insurance; such as social insurance schemes that encourage the cost-ineffective use of expensive care.

We propose to judge the extent of risk protection by the probability (before the fact) or the frequency (after the fact) that individuals will be impoverished by illness, or prevented from obtaining adequate treatment by their lack of income. A financing system does well on this criterion when such events are unlikely at the individual level and hence rare in the population. This measure combines both the size of the risk, and the individual's economic condition. Moreover, to take account of equity concerns, variation in this probability (or frequency) across population groups is a useful indicator for deciding on the priorities for expanding risk protection.

Measuring risk protection, in this way means we do not rely on the public's satisfaction with the available level of risk protection as part of the definition. Instead we believe those reactions will be captured by citizens' levels of satisfaction with the health system as a whole. The effects of insurance arrangements on health status also are not measured by such a concept. Instead we believe such effects will be accounted for by looking at health status directly.

To accurately determine levels of financial risk, defined as we have, requires household survey data on actual health spending and health care seeking behavior. And such studies can be expensive and difficult to conduct – especially in poor countries. But

9

a reformer interested in making a careful diagnosis of national performance on this criterion must be prepared to support such research.

Our three bases for selecting criteria (presented above) help explain why we have chosen financial risk protection as a core performance measure, as opposed to either the overall economic circumstances of citizens or the overall fiscal burden of the health sector. From the perspective of social policy, citizens' net economic circumstances would be the most relevant performance variable – since that is what determines well being (of interest to utilitarians) and opportunity (for liberals). But for the health sector, such a variable does not pass the test of causal impact. Citizens' economic situations are affected by many areas of public policy apart from health sector financing: e.g. tax, policy, public subsidies, directly provided public services, development assistance, infrastructure investment, etc.

From the point of view of general social and economic policy, the health sector's financing burden is an intermediate variable – significant for its impact on the ultimate distribution of well being and not judgeable in isolation from that impact. For example, user fees place a heavier burden, relative to income, on lower-income citizens. But in a society that provides many services to the poor, such fees have a different ethical significance than in a country without such policies. Hence we propose to treat the health sector financing burden as an intermediate variable – as we discuss below.

5. The Role of Cost in Problem Definition

In Chapter 1 we suggested that the worldwide context for health sector reform frequently involves a clash within a country between rising costs and rising expectations, on the one hand, and limited capacity to pay, on the other. Yet this kind of perceived *cost-performance dilemma* has to take account of the international data that show that the same spending yields strikingly different results in different countries[11]. Some nations get good results with much lower levels of spending than others. This pattern implies that while spending more money, can be helpful, it may not be either necessary or sufficient to improve health sector performance. For example, the existing management system and organizational structure in a country might be such that much better performance could be gotten with existing resources. It might even be the case that added funds would be largely wasted, so that the capacity to usefully absorb additional revenue would have to be increased first if more money is to produce better outcomes.

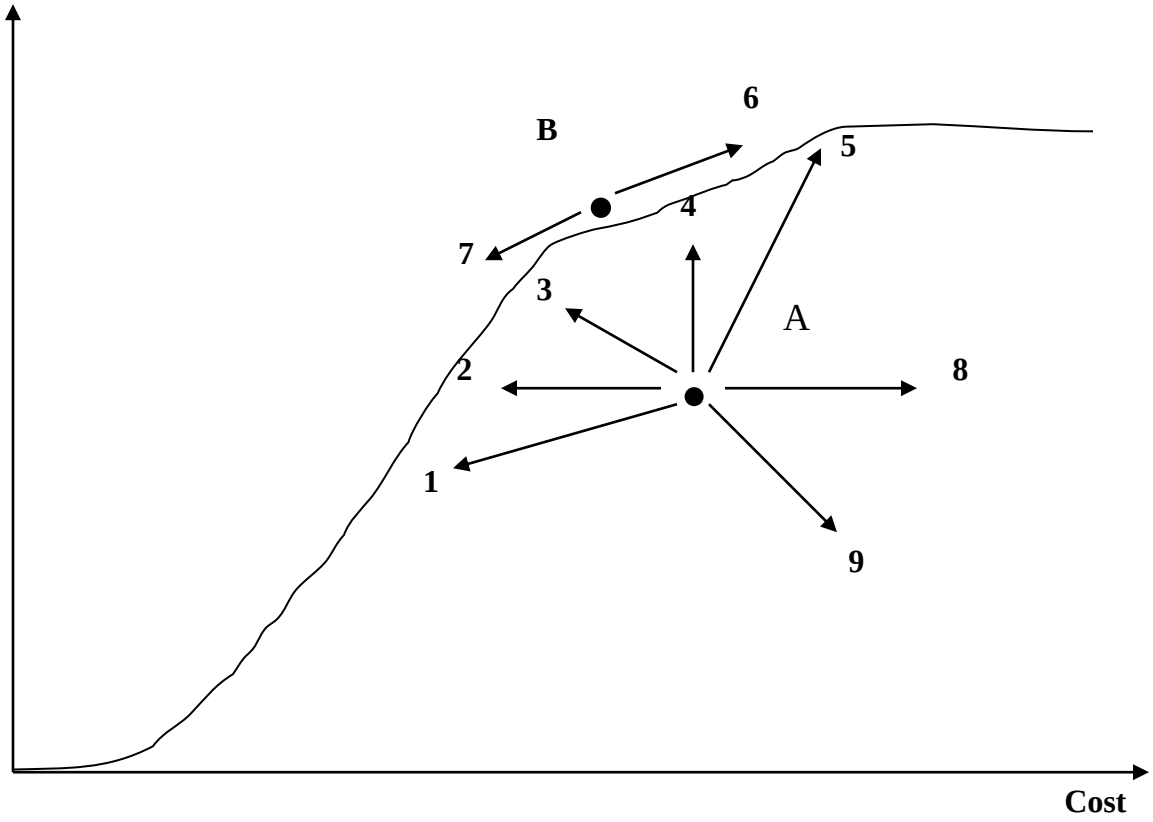
What role does cost play in health sector problem definition? Governments set financing and payment schemes in order to control their costs and meet overall budgetary priorities. The amount available for health may be a residual – determined by the country's economy and by other government programs and expenditures. Hence, in the short run, public funds may be a constraint reformers have to respond to. Thus the question might be, how can we get better performance from existing resources – or even how can we lower health sector costs in response to an economic crisis. On the other

11

hand, financing arrangements are not fixed in stone. So some diagnostic journeys will lead reformers to develop new schemes for mobilizing resources either in general (e.g. broad-based social insurance schemes) or for targeted purposes (e.g. user fees in rural hospitals to generate discretionary funds at the periphery).

The various possible ways cost may enter into a nation's performance problems are thus illustrated in Figure 4.2.

Figure 4.2.
Performance



Since most health systems are not fully efficient, they are in the situation represented by point A (where more performance is possible from current spending). Such nations then face five choices about how to change the relationship between cost and performance:

1) A country is willing to accept somewhat reduced performance in order to significantly reduce cost
2) A country wants to save as much as possible without reducing outcomes
3) A country seeks improved efficiency to both lower cost and raise performance
4) A country is eager to maximize performance for the current budget
5) A country is so eager to improve performance, that it is willing to spend more to do so

In addition Figure 4.4 illustrates the choices open to a nation that is fully efficient (i.e., at point B) – a case we believe is uncommon. In that situation cost and performance have to go up and down together (changes (6) and (7) in the diagram).

In the real world, two more changes can and sometimes do occur. In (8) cost increases without any improved performance. While this may not seem sensible, it can be a response to political pressures to increase employment – and provide patronage. Indeed the cumulative effect of such moves helps explain the substantial over-staffing characteristic of public sector institutions in many countries. Even moves like (9) – when costs go up and performance down – are known to occur. Anecdotal evidence suggests that new regulatory controls in the U.S. Medicare system in recent years have done just that. Doctors have less time for patients, even as costs have risen as more administrators have been hired.

Different players in the health reform debate often do have different views about the nature of the cost-performance dilemma facing their country. Ministries of Finance often argue that the situation is like point A and the needed change therefore is option 3 – more performance *and* lower cost. (In countries going through difficult times due to war or economic crisis, they will even argue for 1 or 2, to get maximum cost reduction.) The Ministry of Health in contrast tends to argue that the system is at B and that move 6 – more spending for more health – is the only appropriate response. Obviously admitting that the country is at A can be difficult for the Health Ministry exactly because it reflects badly on how it is carrying out its responsibilities.

Different aspects of cost are relevant in different contexts. For budget decisions the relevant costs are those paid for by government. Indeed both national and local governments often look only at their own costs, and ignore those incurred at other levels. On the other hand studies of National Health Accounts in many countries have demonstrated that even in poor countries, with extensive public systems, a substantial part of health expenditures comes from patients directly paying both traditional and western practitioners for drugs and clinical services.[10] Hence looking solely at government spending is not sufficient if we are interested in the total burden of the health

10

sector on the society. Indeed the most comprehensive assessment of costs has to include non-money costs, like the time and effort expended by patients and family members. Such an inclusive accounting reflects the economic burden (the *opportunity cost*) of the health sector.[11]

How can a country know whether the cost of their health sector is appropriate? How large a burden should be imposed on the society and the economy to provide health care? That is, how does a country know whether 1, 2, 3, 4 or 5 is the appropriate formulation for its circumstances? One approach is to compare the gain given up by not having the last unit of spending *outside* of the health sector with the last increment of gain from spending more *inside* the health sector. Subjective utilitarians would say that the relevant measuring rod for this analysis is consumer satisfaction. They would explore how much utility was generated by marginal adjustments in spending inside and outside the health sector. In contrast, objective utilitarians have to formulate an index of gain for non-health sector activities (equivalent to DALYs or QALYs in the health sector). Then, they would ask whether the increase in health, which would come from spending more on health, is worth the loss in another area that would result from spending less on some other goal (e.g., environmental protection or education).

Whichever method is used, the broader the scope of the analysis, the more difficult it becomes to compare gains and losses. It is difficult enough to decide how to trade physical pain and suffering against the cognitive distortions of mental illness in order to evaluate shifting budget resources from trauma care to mental health. But can we really decide how to trade either of these against the gains from expanded primary education, or from increased protection for endangered species?

In practice, societies answer such questions incrementally, by changing various spending levels, step-by-step, in search of a reasonable balance.[12] After all, the relevant opportunity cost may not be inside the public budget at all. Rather, it may depend upon what citizens have to give up out of their private spending if taxes, premiums, and/or fees are raised. This multidimensional balancing act is what public sector budgeting seeks to achieve. And as tastes, technology, wealth and external threats all change over time, a society's answer can (and should) be readjusted.

Setting health sector budgets and changing total health expenditure are especially difficult when a country is unable to finance a health care system that will achieve its health status goals. Unfortunately, this is the situation today in many low- to middle-income countries. Simple cost-effectiveness analysis often leads to the conclusions that a country should spend less money on expensive, high technology care and more on basic services. This is because saving the lives of those already old does not produce much in the way of added QALYs—because their life expectancy is quite modest. On the other hand, political pressures to provide such services are often substantial—in part because of the ethical and sociological imperative to "rescue" those threatened with death. In addition, people will spend out-of-pocket care when faced with a serious illness. Hence

11

12

Figure 4.3.

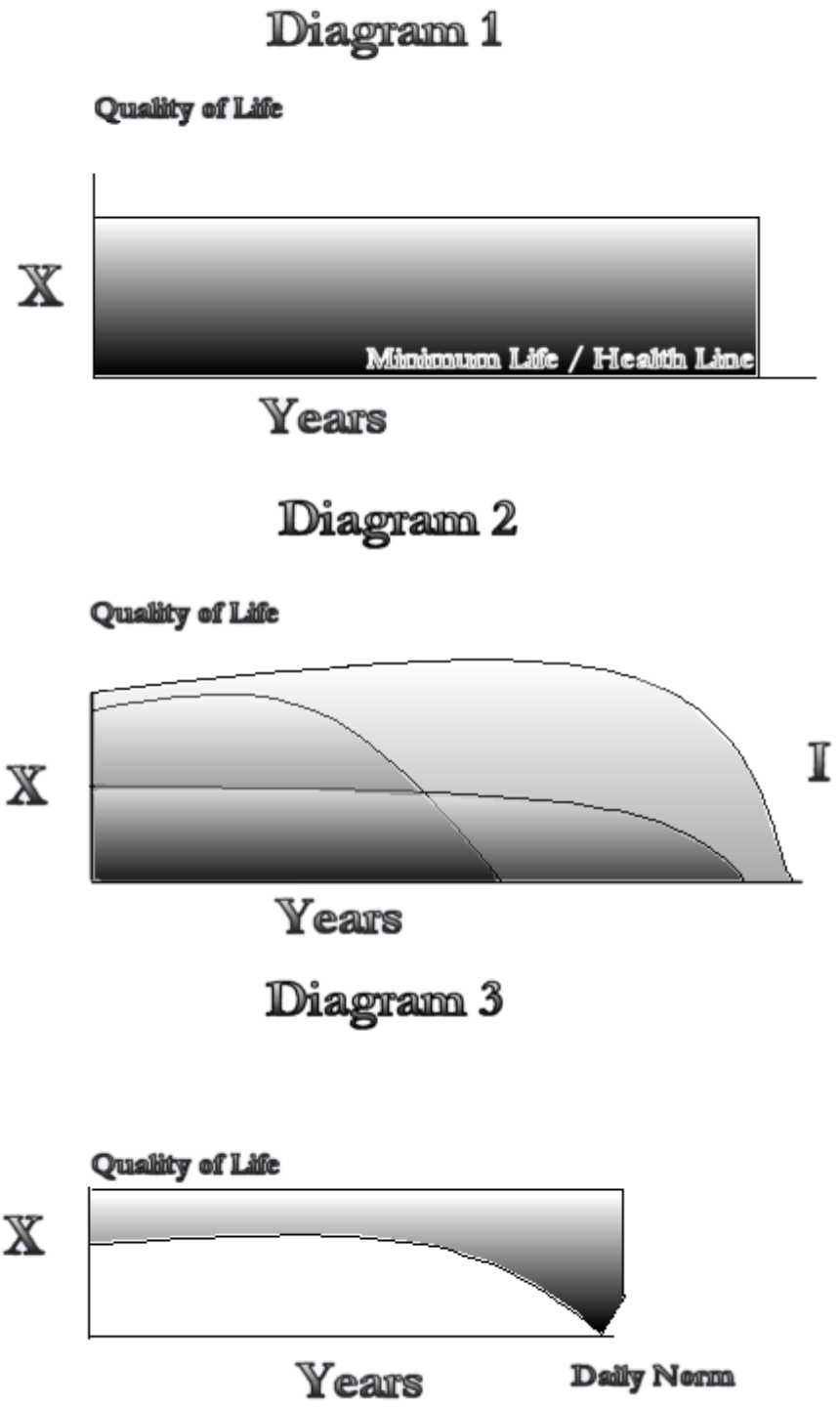


Diagram 1
Quality of Life
X
Minimum Life / Health Line
Years
Diagram 2
Quality of Life
X
I
Years
Diagram 3
Quality of Life
X
Years
Daily Norm

public financing of acute hospital care is typically necessary to achieve risk protection—even if the care itself does not pass cost-effectiveness tests.

Given these conflicts, clarifying alternative values can play a role in shaping a practical compromise. And the agenda for reform often consists of finding ways both to improve performance and raise more money. But a gap may still remain. Here again we believe that responsible reformers need to take the lead in telling citizens the truth about the choices they face and in shaping effective social responses. Not all problems can be solved in a world where resources and knowledge are always limited.

6. Intermediate Criteria

Now we turn to a discussion of our proposed intermediate criteria: efficiency, access, quality, and financial burden. These measures are of interest because of their impact on core health sector performance criteria, or their impact on the general socio-economic system. Our bases for choosing intermediate criteria reflect our methods for choosing core performance criteria: political salience and causal relationship to policy options. But we also need to add impact – the extent to which the proposed measures influence our core criteria. Another way to think of this point is to ask, what features of the system are we liable to encounter as explanations of poor performance, once we begin our diagnostic journey.

6.1. Efficiency

Our first intermediate criterion is efficiency and it relates directly to the question of cost. That is, questions about efficiency often arise when we try to analyze the causes of a particular cost-performance dilemma. In practice efficiency is a widely discussed and politically important concept that economists talk about endlessly and confusingly. We need to disentangle this confusion, in order to clarify how efficiency fits into the process of problem definition and diagnosis.

First we need to explain *technical efficiency*. Economists use this term to refer to situations in which a good or service is produced at minimum cost. An alternative, but equivalent formulation, is that we get the maximum output given our level of spending. For example, is cost-per-day in the hospital as low as possible, or are as many patients as possible being treated for the available budget? Technical efficiency is primarily a concern of managers. It is a matter of *how* we produce something.

In health care, there are thousands of different kinds of services. To be technically efficient, each one must be produced at minimum cost. When this is accomplished, economists describe the result as being on the *production possibility frontier*. This means that more of one output can be produced only by making less of another.

A second notion of efficiency is *allocative efficiency*. This refers to *what* is produced. Have we produced the right collection of outputs to achieve our overall goals? That is, are we at the right point on the production possibility frontier? Since economists are generally subjective utilitarians, they typically use the term "allocatively efficient" to mean the set of outputs that maximizes customer satisfaction. Health planners, when they ask whether a particular set of services produces maximum health status gains, are also asking about allocative efficiency – albeit in terms of a different goal. Improving allocative efficiency is thus the implicit question that confronts anyone who is trying to influence the mix of outputs (e.g., cosmetic surgeries vs. primary care visits) that the health system produces.

Since our framework recognizes multiple conflicting goals, from our perspective evaluating allocative efficiency is inevitably complex. Someone first has to describe the tradeoffs they are willing to make among various objectives. Only then can they know whether they have achieved as much of their goals as possible.

Countries that confront a cost-performance squeeze often seek ways to improve both kinds of efficiency through the process of health sector reform. From a political perspective, however, it is often more difficult to improve allocative efficiency then technical efficiency. Lowering the cost of producing outputs (to raise technical efficiency) is not easy, because workers often resist change and effort is required to reorganize production. Still it is a relatively straightforward managerial task. In contrast, improving allocative efficiency imposes very high costs on those involved in the production of those outputs being scaled back and on the consumers of those services. Shifting resources (people and money) among activities is thus often vigorously resisted.

Before moving on we also need to clarify the confusing claim that efficiency and equity are either necessarily – or at least frequently – in conflict. This argument is based on the premise that the term "efficiency" should be used very narrowly; for example, to mean maximizing average health status without regard to distribution. In that context, serving those in poor health in rural areas can be characterized as "inefficient" because more DALYs or QALYs could be produced for the same money if it were not spent with an eye to equity. But there is no reason to restrict the definition of "efficiency" to such a narrow context. In our framework the "efficiency" of the system (both allocative and technical) is determined by whether or not it reaches society's goals at minimum cost. Thus one could sensibly ask – is a nation's health system efficient in reaching its equity goals (e.g., does it provide health status gains in rural areas at minimum cost)?

In summary, both kinds of efficiency refer to the *relationship of inputs to desired results*. Technical efficiency means producing outputs in the "right way" – at minimum cost. Allocative efficiency means producing the "right outputs" – to maximize the achievement of our goals. Unless a health care system is *both* technically and allocatively efficient, it will not achieve as much as it might – with a given budget. Efficiency then is clearly an intermediate criterion. It is not an end in itself, but rather a condition for achieving one's ultimate ends – whatever they are. To maximize health status, or consumer satisfaction, or any other goal, the health system has to be efficient. As health

reformers proceed on their diagnostic journey, they are sure to confront the lack of efficiency as a potential cause of inadequate health sector performance.

6.2. Access

In discussions of health sector reform, *access* is often a major concern. A lack of access is often introduced as an explanation of poor health status in rural areas, or low levels of satisfaction among the poor. Yet in these discussions, the term is often used to refer to different concepts. First, access sometimes simply refers to whether services are offered in a specific area. Here, the question is *physical availability*, and it can be measured by the distribution of available inputs (beds, doctors, or nurses) compared to the population. A second notion, one that more closely reflects the intuitive meaning of the term, is *effective availability*: i.e., is it easy for citizens to get care? Differences between physical availability and effective availability can arise because various barriers (cost, travel times, poor service) may keep people from using facilities that are physically available. However, effective availability is difficult to measure. It is not easy to collect the data on prices, service levels, waiting times, and cultural acceptability, that would be required to construct a valid index of effective availability.

In practice, the term access is often employed to refer to *utilization*. Per capita measures like hospital admissions or outpatient visits are computed for various population groups and those with low use are said to lack access. But use is only partially a reflection of effective availability – since patients may choose to not use services, even if they are available. If we ignore the possibility that low use might reflect patient choices and instead argue that low use always means that there are some barriers to care, access has been abolished as a meaningful concept since it is no longer independently measurable apart from utilization.

At first glance, access defined as effective availability is an obvious intermediate criterion. It does influence both health status and consumer satisfaction. And not considering it a core criterion is ethically quite in keeping with utilitarian perspectives – since utilitarians view availability as a means and focus on either health or satisfaction as the end. Similarly, egalitarian liberals, who focus on health as an aspect of opportunity, are also likely to view access as a means to providing the minimum quality of life they see as a positive right.

There are, however, some who do argue for using access as a core criterion. Some egalitarians (like Amartya Sen) argue that government is obligated to make services available, and then let citizens use these (or not) as they choose. In this view, the effective availability of health care (not health itself) would be a core goal. Similarly, some communitarians focus on the distribution of health services as a matter of fairness to various local communities. From that perspective access, defined as physical availability, is often a focus of intense political discussion. Indeed in advanced countries, changes in effective availability (and any associated impact on customer satisfaction) may dominate political discussions since marginal changes in already high service levels often have little observable impact on health status.

On balance, however, we believe that access fits better as an intermediate rather than a core criterion for health system performance. First, as long as the health system is considered a means, effective availability also is a means, not an end. Second, our framework has significant flexibility. For example, providing services to increase satisfaction, even if it does not increase health, is a possibility we take account of within our framework. (One reason for doing so, for example, might be to respond to what economists call "option demand" – the value to citizens of having something available even if it is not used.)[13] Where physical availability is valued apart from its consequences, we believe such concerns can best be understood as community norms. These norms, in turn, can be seen as constraints upon the process of health sector reform – as we discuss below – and therefore need not be given the status of a core criteria.

Viewing access (i.e. effective availability) as a tool for improving health status and satisfaction has considerable implications. It means, for example, that we should view with skepticism arguments that every city ought to have certain services – because such a situation is "fair"—if such facilities are not part of a cost-effective plan to increase either satisfaction or health status. Such arguments can only prevail if we conclude that they reflect important communitarian commitments in a particular society. Moreover, if services are valuable only if they produce outcomes, the reverse is true as well. The lack of services is most significant when outcomes are unsatisfactory. We realize that actual debates over health reform are rarely conducted this way. But we believe that particularly in poorer countries with quite limited resources, seeing access primarily as a means places a useful argumentative burden on those who contend otherwise. We believe this way of thinking promotes greater analytical rigor in conversations about priorities for health sector reform.

6.3. Quality

Quality is our third intermediate criterion, valuable not for itself, but for its role in achieving core performance objectives. Poor quality (like poor access) is often invoked as an explanation for performance failures in many different contexts. Advocates of different reform agendas try to appropriate the term to support their goals. Thus "quality" is sometimes defined from the patient's point of view and sometimes from the perspective of the doctor. It is sometimes applied to the treatment of a particular case, or to the care provided by a particular hospital, or comprehensively to a national system as a whole. In general we would argue it is most meaningful to speak of quality at a very disaggregated level – the treatment of a particular patient in a given encounter. More encompassing judgments (of particular doctors, hospitals or national systems) reflect aggregations (averages) of such encounter-level experience.

Even at the encounter level, the term quality, like the term access, refers to several different phenomena (see Table 4.2). To explain the relationship between quality and our core goals we first have to distinguish among the various meanings of the term. In what follows we abstract from *who* is doing the judging to try to identify various measurable

13

quality aspects. But as we stress repeatedly, different individuals (e.g., different doctors, patients, politicians) may place different importance on different quality dimensions.

Table 4.2

Meanings for the Term "Quality" in Health Care

Quantity	Clinical Quality	Service Quality
	• Human Inputs: skill, decision making	• Hotel Services • Convenience: Travel and waiting times
	• Non Human Inputs: type & quantity, Production System	• Interpersonal Relations: Care, concern, politeness • Respect: Information, decision making

First, as the table indicates, "quality" can be used to mean simply the *quantity* of care provided to a patient, as in: "My aunt got the highest quality care. They did everything for her." Americans who praise our system for its high "quality" often are using the term in this way. Second, health professionals typically use quality to refer to *clinical quality*. With regard to human inputs this involves both the skill of caregivers (e.g., the surgeon's technique), and whether or not correct decisions are made about diagnosis and treatment. On the non-human side, it involves whether the right drugs and equipment are available. Quality is also a product of the effectiveness of the system of production that combines these inputs into actual delivered services. Third, patients, who generally find it difficult to judge clinical quality, tend to focus on various aspects of *service quality*, which they can observe and evaluate. This involves several subcategories including hotel-type services: matters of convenience, how people treat you interpersonally, whether they treat you with respect.

Indeed each of these aspects of service quality are themselves multi-dimensional. Hotel services include food, cleanliness, the nature of hospital and waiting rooms, etc. Convenience includes: travel times, waiting times, opening hours, time to get an appointment, etc. The interpersonal dimension involves whether providers are polite and emotionally supportive. Finally are patients given appropriate information and treated with respect? (There are complex issues about the role of patients in the care process – issues to which we return to below).

To understand the role of quality, as an intermediate performance goal, we have to take account of these multiple meanings. Just as a car has various quality dimensions (fuel economy, acceleration, passenger capacity, etc.), so too does health care. Increased quality (on any dimension), in cars or care, might not be worth the cost – depending on the values of the person making the judgement. And having more of one quality (e.g.,

luggage capacity or patient choice) might lead to less of another quality (handling or correct clinical decision-making).

Measuring quality generally requires detailed and expensive data. Service quality can be measured in various ways. Good administrative systems can generate data on parameters like waiting times, delays in getting an appointment, etc. Special purpose quality monitoring systems can also be created (e.g., periodic inspections of cleanliness or expert evaluations of food). Inter-personal relationships can be evaluated using patient reports. These can also be useful in identifying trouble spots – from the patient perspective – in aspects of service (e.g., food and cleanliness). Again – as in the case of satisfaction scales – industrialized countries have invested heavily in developing the relevant measurement tools.

Clinical quality can be evaluated by comparing clinical care – as recorded in patient records for example – with expert opinion as to appropriate treatment. It can also be measured indirectly by outcome data like infection rates, operative mortality, etc. Data on the adequacy of production systems – which is the focus of the whole Total Quality Management approach – require even more detailed institution studies. The difficulties of collecting and interpreting such data, help explain why many countries rely heavily on regulating *inputs* (e.g. educational requirements) rather than monitoring and evaluating processes outcomes, in the quality arena. (We discuss this point further in the regulation chapter below.) Indeed monitoring input quality (e.g. does the health center have needed drugs and equipment, is a doctor available in an emergency) often is the only – albeit imperfect – way patients have of monitoring clinical quality.

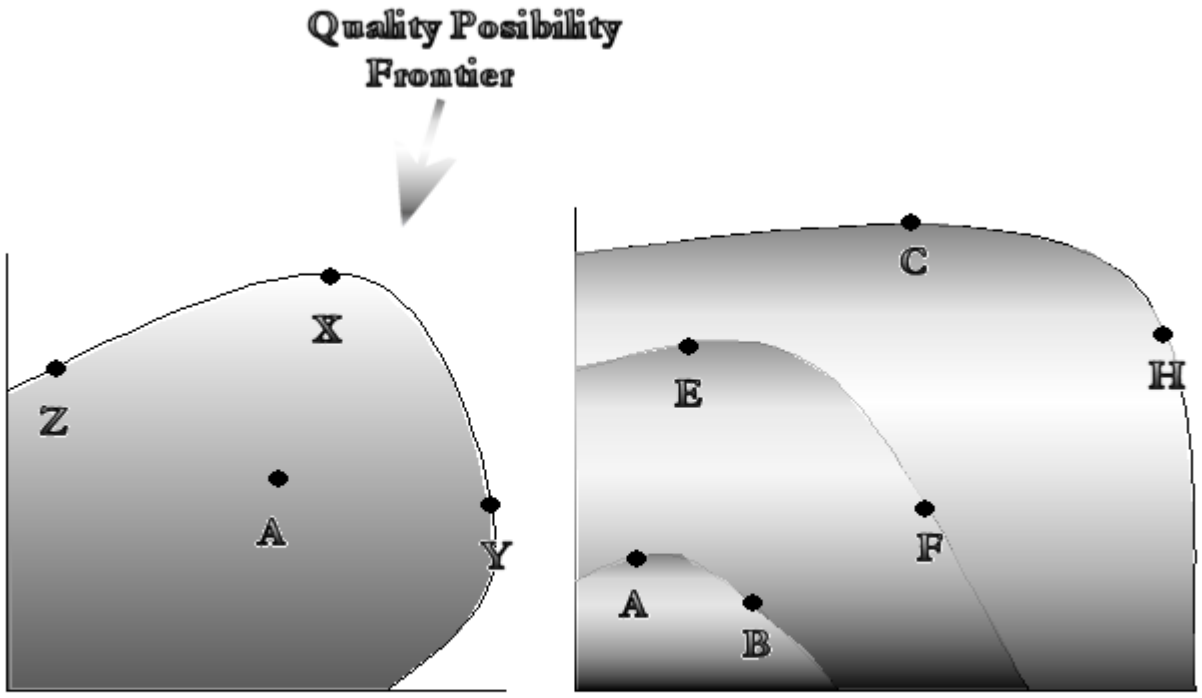
Not just the average, but rather the distribution of quality – i.e. who is subject to poor quality – is often important both politically and practically. For example, suppose a country is concerned with patients bypassing local health posts to seek treatment at regional or national centers. Deciding what to do about such a situation requires an analysis of both clinical and service quality – as well as the quantity of services available – at the local level. If quality at the periphery in any or all of these aspects is poor, it will be more difficult to prevent widespread "by passing." And allowing such poor quality to persist is likely to particularly disadvantage the poor, who cannot afford the time or expense of traveling further afield for care. As with other performance measures, summary statistics on equity are much less helpful to a reformer than data about the distribution of poor experiences by site or patient group.

When it comes to choosing what quality to produce, a health care system has three tasks. The first of these is simply a matter of insuring a particular kind of technical efficiency. Is every service produced in a way that results in the highest possible quality given the costs being incurred? Having less than the maximum attainable quality, for a given cost, is to be technically inefficient in the production of quality.

To see this, consider Diagram 1 of Figure 4.4, which depicts the "quality possibility frontier" for a particular health care service. To make a diagram possible we have made the simplifying assumption that "clinical" and "service" quality are each a

single distinct magnitude (we get to the issue of "quantity" shortly). In fact the “frontier” really exists in a space of many dimensions. For a specified budget, any service can only produce a limited amount of clinical or service quality. We have drawn the diagram to imply that if either service or clinical quality is too low, the other aspect of quality suffers; but that is not necessary to the argument. What is key is that to be technically efficient in the production of quality a service has to be organized in a way so that its output is *on* the frontier – not at some point like “A.”

Figure 4.4



However, once they are on the quality possibility frontier, managers still have to decide what *mix* of qualities to offer for a given budget. (This is like the question of what kind of car to design.) In terms of the frontier, should they be at X or Y, or somewhere in between? (Given the way we have drawn the frontier, there is no reason to be at Z since it is possible to do better on both dimensions.)

But there is yet a third quality task. As shown in Diagram 2, there are a *series* of quality possibility frontiers for any given service: each based on a different *budget* (I., II., and III. in the diagram). The third task then is setting the *level* of spending for each service – which in turn determines just what mix of quality levels can be produced for that service. This is the quantity issue we identified initially – which really amounts to determining the available level of resources.

How would different reformers, with different philosophical views, perform these tasks? Objective utilitarians, interested in health maximization, would want to produce the maximum level of clinical quality for any given budget. That is, they would always pick a point like X in Diagram 1. They would then choose a budget for each service on the basis of marginal cost effectiveness analysis – spending money on different services in a way designed to get the biggest health status gain.

A subjective utilitarian, interested in maximizing customer satisfaction, faces a more difficult decision. Patients typically care about many aspects of both clinical and service quality. Thus, for each service, subjective utilitarians have to decide on both what mix of qualities to provide and on the expenditure level – and do this in a way that best responds to the varied individual preferences in the society. The obvious difficulty of such a task helps explain why so many subjective utilitarians favor using markets in health care (and everywhere else). Markets allow everyone to choose (and pay for) the set of services, and the mix and level of qualities for those services, that they prefer. Of course this assumes that customers (i.e. patients) can judge quality levels – which can be doubtful when it comes to clinical quality. Hence even some subjective utilitarians would regulate markets for quality purposes – as we will discuss in Chapter 9.

For liberals, who believe that autonomy is a basic right, one particular aspect of quality assumes special importance: Liberals want to insure that patients are fully informed, and that they have the power to determine their own care in-so-far-as they want to exercise that right.[14] Relativist communitarians, in contrast, accept the fact that in some cultures treating patients with liberal respect – by for example insisting on informed consent before treatment, is not the norm. Our solution to such disagreements is to say that higher quality means allowing patients to play whatever role they want to play in their own care. Thus they can choose to not choose, when that is consistent with their cultural perspective. (We return to this issue when we discuss community values below.)

In summary, improving the management of quality is a potentially important aspect of health sector reform – exactly because quality is an important intermediate variable affecting both health status and customer satisfaction. We have identified three tasks in that context. First, is the system technically efficient in the production of quality so that quality (on various dimensions) is as high as possible given the budget? Second, does the system produce the right mix of qualities given budget levels for each service? Third, is the question of quantity. That is are budget levels set to produce appropriate levels of quality for each service? The day to day responsibility for getting to the quality possibility frontier, as well as for determining the mix of qualities, rests with the

14

managers of health care services. And as we will see, when discussing organization, how those managers are chosen and rewarded will influence their decisions. However financing and payment systems and regulatory regimes, determine the budgets and incentives confronted by providers – and hence influence their quality and quantity choices. And these latter decisions are made at the health systems level. In short, quality is everybody's business in the health sector – once we understand the complexities this deceptively simple idea actually involves.

6.4. Financial Burden

Our fourth intermediate criterion has to do with the burden of the financing system. We explained earlier why this is an intermediate criterion – of interest for its impact on other variables – as opposed to a core performance criterion. Simply put by financial burden we mean who actually bears the cost of the health systems. That is, whose economic situation is changed (and by how much)? The actual burden may not be in the form of making a direct payment but rather in the form of some other change – like higher prices or lower wages.

Most usually the distribution of financial burdens is analyzed in terms of the impact on different income groups. The difficulty in determining that distribution is that the person who directly pays a tax may not ultimately bears the burden of it. For example, an employer's contribution to a social insurance system is likely to result in some combination of lower wages to employees, lower profits to owners and higher prices to customers. Determining the burden of a particular tax in a particular country will thus generally require sophisticated statistical research, to trace out the eventual effects. For example, the effect of different tax rates in different industries can be studied to see how prices and profits are effected, and those estimates then applied to, for example payroll taxes used to support healthcare.[15]

There is an additional complication. Sometimes an increase in government health spending is not financed by tax increases, but by cuts in government spending elsewhere. In that case the new programs are being paid for by those who lost benefits due to such budgetary reallocations. Moreover in many countries some or all health expenditures come from the general government budget that uses many sources of revenue. How can we know which programs are smaller than they would be, or taxes higher than they would be, if health expenditures were less of a burden? Yet such hypotheses have to be answered if we are to determine overall "burden".

The economist's standard answer to this question is the concept of "balanced budget incidence." This procedure tells us to estimate what taxes are likely to go up, or expenditures go down, to accommodate any new spending patterns. Such a method is best adapted to exploring the *marginal* effect of adjustments in spending – rather than to giving an answer about the overall distribution of the total burden of the entire current level of public spending on health. That is, it is seldom possible to answer the question, "What would all prices, incomes and tax burdens be like in the economy if the

15

government did away with *all* health spending?" For health system reformers, however, the answer to such a question is seldom important. Instead it is precisely more tractable marginal questions which are relevant since they involve discovering the effect of some proposed change in the financing system. Yet, we would argue that even such marginal impacts are only an intermediate variable – of interest because of their ultimate impact on individual well being.

Once the change in burden that would be produced by a proposed change in health sector finance has been determined, the next question is, how to judge that impact. This is clearly a matter of ethics and values. Economists focus on the distribution of burdens across various income groups – a concept they call "vertical equity." For complex historical reasons the critical reference point for such an analysis is a tax that takes the same *percentage* of income from the poor as the rich. Such a tax is called "proportional." A tax which falls more heavily on the rich (i.e., one which takes a higher proportion of their income) is called "progressive." One that takes a higher percentage of income from the poor is called "regressive." The underlying philosophical analysis here is utilitarianism – but the particular argument is now so little known that many economists don't know of the origin of or justification for their concepts.[16]

Depending on its values, that is on how "pro poor" a country wants to be, it can design a system of taxes, fees, and premiums that best meets its goals. User fees, and insurance premiums that are not income-based, are obviously the most regressive. Since they take the same amount of money from rich and poor, they take a much larger percentage of income from the poor. Progressive taxes on all forms of income (including income from investments) potentially puts the largest relative burden on the rich – depending on actual tax rates and collection practices. Social insurance schemes, financed by payroll taxes, are slightly regressive. Earned income (as opposed to investment income) is a greater percentage of all income at lower income levels. Hence a tax on earned income takes a higher percentage of total income from those at lower income levels. This is especially so if there is an upper limit on the total payroll taxes paid by an individual – as is often the case. Sales or value-added taxes also generally have a modestly regressive impact, (depending on what, if any, goods are exempt from taxation) since consumption is a higher percentage of income at lower income levels.

Egalitarian liberals look at the distribution of financial burden somewhat differently. For them, a minimum level of economic resources is needed if a family is to have a basic level of opportunity. It follows from such an analysis that those who might be pushed below such a minimum level should pay little or nothing for health care. Instead egalitarian liberals favor income redistribution from the rich to the poor – for that would lead to a reasonable range of choice for all. It is exactly such concerns that lead some egalitarian liberals to oppose user fees for health care services – unless some provision is made to exempt the poor from such burdens.

Economists also analyze financial burden in terms of "horizontal equity." This concept says that those at the same economic level should be treated the same. In the

16

health reform context, the largest issues in this regard tend to be regional. This particularly occurs when, as part of decentralization, a country shifts the financing burden for health care from the national to local or regional levels. Poor areas then need to pay higher taxes if they are to provide the same level of service as richer areas. Similar questions arise between rural and urban areas. Since the cost of providing services to scattered rural communities is typically higher, local financing means that rural areas must either pay more or get less. These inter-regional variations in tax rates, economists argue, violate the principle of horizontal equity. Certain other taxes – on say cigarettes or alcohol – can also raise horizontal equity issues. They not only burden selected consumers, but also can have a disproportionate impact on farmers, workers, and businessmen who are part of the taxed industries.

7. The Relationship Between Intermediate and Final Criteria

Since we have described three final and four intermediate criteria the question naturally arises, how do these relate to each other? This is an important question in part because it sometimes is easier to relate the actions of the control knobs to the intermediate rather than the final criteria. More generally, an overview of these relationships allows us a first step down the path of diagnosis described in Chapter 6.

7.1. Poor Health Status

If some or all of the population has poor health status in general we are looking at one of four causes (1) not enough services are available; (2) the services are not producing the desired results; (3) the services are not being used; (4) The wrong kind of services are available. How do these reflect our intermediary criteria?

If not enough services are available (number (1) above) this has to reflect either technical inefficiency (high unit cost) or that not enough resources are being devoted to the activity. Technical inefficiency in the production of clinical quality is also behind cause 2 – not producing potentially achievable health status results given the available resources. Services not being used is a matter of a lack of effective availability and/or customer response to the available mix of clinical and service quality. And wrong output mix (cause 4) is by definition a matter of allocative efficiency.

Consider a specific example. Suppose a country has poor health status in remote rural areas – where there are few services because the nation feels that given the limited health budget it cannot afford the high cost of expanding its system of health costs to such areas. Moreover where posts have been established they are often undermanned, undersupplied and under used by citizens.

This situation in part reflects technical inefficiency – the system of delivery care to those areas has to be made less costly. There is also perhaps allocative inefficiency – the curative services the state is struggling to provide may not be the most cost effective and perhaps it should focus more on prevention. So it is both doing the wrong things and

doing them badly. The lack of supplies means that the posts that do exist are inefficient in the production of clinical quality. The lack of utilization by citizens may reflect their dislike of the quality mix being offered. Also there may be various barriers (e.g. user fees) that prevent physical availability from producing effective availability – thus limiting access.

Note that *not* all potentially relevant causes are under the headings of our intermediate criteria. Resource limits – fiscal or material – are one such example. So too are citizens' attitudes. But the performance of the health sector in the intermediary criteria is an important part of the story.

7.2. Citizen Satisfaction

When we ask about causes of potential citizen dissatisfaction we again find our intermediary criteria helpful. (1) Services may not be available; either physically unavailable or effectively unavailable do to various barriers. (2) The services that are available may not be to the liking of citizens – either because overall levels are low or because the quality mix does not suit their tastes.

This leads us to an interesting point – namely the intermediary performance criteria can be casually linked to each other as well as to the final criteria. Thus a lack of physical availability (an access issue) can arise if costs are too high (a matter of technical efficiency).

Again our concern as reformers is likely to not only focus on the average. Instead areas where, or groups for which performance is especially poor are likely to be a particular focus of concern. This again shows why detailed data – rather than summary statistics – are especially relevant to priority setting and diagnosis.

7.3. Risk Protection

Suppose we do find that some citizens are impoverished by the need to pay for care, or that they have poor health status because they can not pay for physically available services. Then what we have is clearly a problem of effective availability – or the lack there of – due to price barriers. Do note that supposedly free services in the public system are often not really free when one adds up all the costs drugs and supplies paid for by families, gratuities or under-the-table payments to staff, the income lost from time devoted to nursing care.

However where affordable public services are not utilized, it is likely to be a matter of low quality – due to inefficiency, low budgets, and/or the lack if service "qualities" that are available. So again we see the intermediary variables are casually linked to each other and to the final variables. Indeed even the final criteria variables can be inter-connected since a lack of risk protection can produce poor health status.

Now in none of these cases are the intermediate criteria themselves the end of the story. The question immediately arises, why is technical quality low or physical availability absent or service quality poor or whatever. There are additional stages on the diagnostic journey to go through again as we discuss in Chapter 6. Our point here however is that the intermediate criteria do provide a handy checklist of causes to consider.

8. The Cultural Context and Community Values

How do we deal with the inter-relationship between health sector reform and the values held dear by a specific community? A surprising variety of cultural rules, taboos and customs do impact the health sector. The list is far too long to present inclusively but here are some relevant examples. One obvious area involves beliefs related to issues of life and death: including abortion, resuscitation, and physician assisted suicide. A second set of issues relates to the control of various substances – from tobacco and alcohol to caffeine and cocaine. The third group of issues involves sexuality – from prohibitions aimed at homosexual or premarital sex, to views on prostitution and contraception. Many societies also have views about how the body itself (alive or dead) should be treated. These shape the acceptability of practices ranging from transfusion and transplantation to customs about male circumcision and female genital surgery. There are also social dynamics that effect health status – ranging from intra-familial patterns of food distribution (e.g. adult males eat first) to gender based variations in medical care utilization. Social norms about the role of western biomedicine also impact the system in many countries. The list could go on and on: the role of traditional birth attendants, whether men and women can be treated by doctors of the opposite sex, the willingness of patients to talk openly to doctors, etc., etc. Clearly in evaluating any proposed policy, the extent to which it would violate such norms, and what that would mean for political feasibility, implementability, etc., needs to be considered.

How seriously a reformer is willing to respect such claims does depend on their ethical perspective. For relativist communitarians, local customs and values should be central to defining any country's objectives for its health sector. For liberals, in so far as individuals want to make certain choices dictated by custom, that is fine. But it is the individual choice, not the community norm, that is worthy of respect. Hence community norms that restrict freedom or opportunity (e.g. no education for girls) can be opposed by egalitarian liberals.

For subjective utilitarians, if observing custom increases someone's utility then that individual should be free to do so. But again, unlike communitarians, and like liberals, for subjective utilitarians custom is a means, not an end in itself. Objective utilitarians are likely to be even more goal directed – respecting local custom only in-so-far-as doing so increases health status (e.g. by facilitating increased patient compliance).

How a reformer wants to treat community norms in the process of policy design, is yet one more decision that has to be made in the health sector reform process.

Reformers who are not relativist communitarians will not see preserving local customs as an important goal, especially if such customs undermine the reformer's core goals. Yet a wise reformer will take local custom into account (if only out of prudence) in thinking about how to proceed. Custom can limit both political feasibility and implementability or both. Hard work can sometimes get even controversial policies adopted politically, and creative social marketing efforts (as we discuss in Chapter 11 below) can change attitudes. Thus customary constraints are often somewhat elastic. But they always exist and always need to be considered.

One set of "customs" we want to call special attention to are current western views about patient choice and informed consent. These customs call for patients to make all crucial treatment decisions, and to do so on the basis of complete disclosure by the doctor of all information about the patient's diagnosis and the consequences of different ways of proceeding. Advocates of the universal applicability of such norms make standard liberal arguments. Treating patients in this way, they say, respects patients' fundamental human rights of autonomy. Relative communitarians, in contrast see these norms – like any other norm – as a particular cultural practice, and believe they should be followed only in those countries where they are widely accepted. By the same token, they argue different traditions or practices also deserve respect when they are widely accepted in other cultural contexts. In many East Asian societies, for example, elderly patients are seldom told of a fatal diagnosis nor would it occur to either such patients or their doctors to have those patients make all the critical decisions about their care.

Our view, as noted above, represents a bit of a compromise. We are sufficiently liberal to suggest that patient choice should be respected either where that is the custom in the society or where a particular patient – contrary to local custom – wants to exercise such choice. The failure to do so is likely to be perceived by patients, family, etc. as a satisfaction-diminishing deficiency in the service quality dimension we labeled "respect" in Table 4.2 above. Where neither of these conditions is met, and both the society and individuals are comfortable with other ways of proceeding, we would not substitute our judgement or standards for theirs.

In summary, like the intermediate criteria we've discussed, local customs are likely to effect health systems performance. Hence they have to be taken into account in accessing the political feasibility and the implementability of policy proposals. Moreover, a reformer's philosophy is likely to play a role in the extent to which they try to alter or avoid such customary barriers – or treat them as worthy of inherent respect. Again, what we urge is explicitness and transparency in how such decisions are reached and explained to others.

9. Developing a Strategic Performance Problem Focus

We have tried throughout this chapter to not just elevate our particular vision of the good society into a universally applicable statement of the goals of the health sector. Instead we have tried to present a framework for defining problems that various reformers can use and adapt to their particular circumstances. That framework involves three core performance criteria, four intermediate criteria and two ethical tests (average and distribution). It also explicitly considers the role both cost and culture should play in the analysis. Hence, it is not a simple analytical structure. However defining problems and evaluating alternative solutions in the arena of health sector reform is a difficult and complex task. Ignoring that complexity achieves a false simplicity – one that trades rhetoric for relevance, slogans for sophistication.

Having said all that, we would like to offer some advice to would-be reformers about what performance problems to focus on, advice that draws together the various threads of this chapter. The essence of our advice is this – think strategically! Choosing a problem definition is the first step in the long process of health sector reform. Hence it should be done with an eye on the larger implications of that decision.

As discussed in Chapter 1– the process of reform is often provoked by events outside a reformer's control. The economy turns down, expectations and/or health costs go up, political change occurs, etc. What we are urging is that reformers make an effort to seize the initiative; to advance their own problem definition in order to shape and orient public and political debate.

Strategic analysis focuses on finding ways to accomplish ones goals through systemic thinking. What then does it mean to be strategic about problem definition in health sector reform? The goals that have to be considered in such an analysis actually have several components. First most individuals have goals for the society itself – that is the kind of thinking involved in the philosophical analysis we have discussed. But real reformers also care about consequences for themselves and their friends – to their political party or their professional group, their hospital or their minister, etc. These various consequences depend not only on what problems are chosen but also on what policies are selected and on what results those policies produce. Short run political gains may come from simply trying to solve a problem. But long run improvements in outcomes, and credit for producing such improvements, depend on whether or not the "try" is well designed. To clarify such issues for themselves, we suggest reformers ask themselves three questions; questions that seek to synthesize the ethical, political and substantive analyses we believe are behind selecting any particular health sector priorities.

- What improvements in health sector performance are most important ethically?
- What areas of poor performance can you reasonably hope to do something about?
- What will the political consequences be of taking on this issue?

Paying special attention to the first question involves giving priority to one's own value commitments. The second question requires consideration of whether there are any promising policies or programs available for dealing with ethically important issues, and whether such measures are likely to be politically feasible and implementable. For only then can a reformer decide if focusing on a particular problem will produce useful results. The cost of such initiatives – compared to the available funds – is a key part of such an analysis. The third question suggests a look at political fallout to round out the analysis. In a sense politics and feasibility can be viewed as "screens" or "constraints" on a value based approach. *Start with what you want to accomplish and then see if there is a way to make that happen at acceptable political cost.*

As Figure 4.1 illustrated, the control knobs are *not* the only forces that determine the performance of the health sector. Instead non-manipulable factors may be at the heart of a particular performance problem. In addition sometimes it will not be easy to know if an effective, feasible and implementable policy option exists without doing extensive research into a particular problem area. But the advice we are giving is – consider that question, as best you can, before you get too far down the road to reform.

We are perfectly well aware that the answers to these questions may *not* all point in the same direction. Sometimes leaders focus on problems where it is unlikely they will be able to accomplish much because doing so satisfies certain political concerns. The U.S. "War on Drugs" comes to mind in that regard. We are not saying such symbolic actions are always wrong or inappropriate. Rather our point is that reformers need to be aware of their own motives and options in such situations.

These questions also remind us that the process of strategic problem definition has to be "agent relative." That is, the political consequences of a given decision are likely to be quite different for different actors. The Minister of Finance may well have a different view from the Minister of Health about the most important goals of health sector reform – and suffer different consequences if, for example the new social insurance fund becomes insolvent. Similarly, focussing on the poor health status of the rural poor is more likely to be politically advantageous to parties trying to attract the support of small farmers – and less helpful to parties whose political priorities preclude developing a significant rural constituency. Thus there is no way to pick a problem definition from an abstract point of view – what philosophers call the "view from nowhere." Instead choosing a particular problem definition has to be seen as a response to a particular individual's or group's circumstances.

When we say reformers have to decide on a problem definition, we also want to call attention to the fact that they will often also have a question of *scope* to consider. Some reformers, in some countries, might choose to focus quite narrowly on one or two specific performance parameters (like high infant and maternal mortality rates in poor rural areas). Such a problem definition is likely to lead to a relatively targeted set of reforms – e.g. the development of new reimbursement mechanisms or selective investment in certain facilities or training programs, etc. On the other hand broader problem definitions are likely to lead to broader and more complicated reform agendas.

Concern simultaneously about widespread failures of risk protection, popular dissatisfaction with the health care system and high costs could lead reformers to a much more ambitious reform program. That is the country might decide to create a new social insurance fund, new payment outcomes for doctors and hospitals, and new forms of hospital organization – all at once. Clearly in making decisions about the scope of problems to tackle, reformers would be well advised to think carefully about the questions of ethical priority, likely effectiveness and political feasibility we highlighted above.

As a practical matter, one way for a reformer to proceed is through "benchmarking" – that is, comparing national performance with various standards to see where performance is both inadequate and potentially improvable. This process can take many forms.

- *Ethical benchmarking* : comparing performance to general ethical norms
- *Internal benchmarking* : comparing performance across groups or regions within the country
- *Historical benchmarking* : comparing performance to a nations' own prior performance
- *External benchmarking* : comparing performance with that of other, similarly situated countries

The first two kinds of comparisons may help answer the first question above – about ethical importance. If a situation directly violates an ethical norm or if a failure is important philosophically that starts us on our way. Internal variation also immediately raises the possibility of equity concerns – if our ethics make that a priority. Internal historical and external benchmarking can help answer the second question: can anything be done about the problem? If we once did better, or if other countries like us do better, or if we do better in some places – all that suggests improvement ought to be possible.

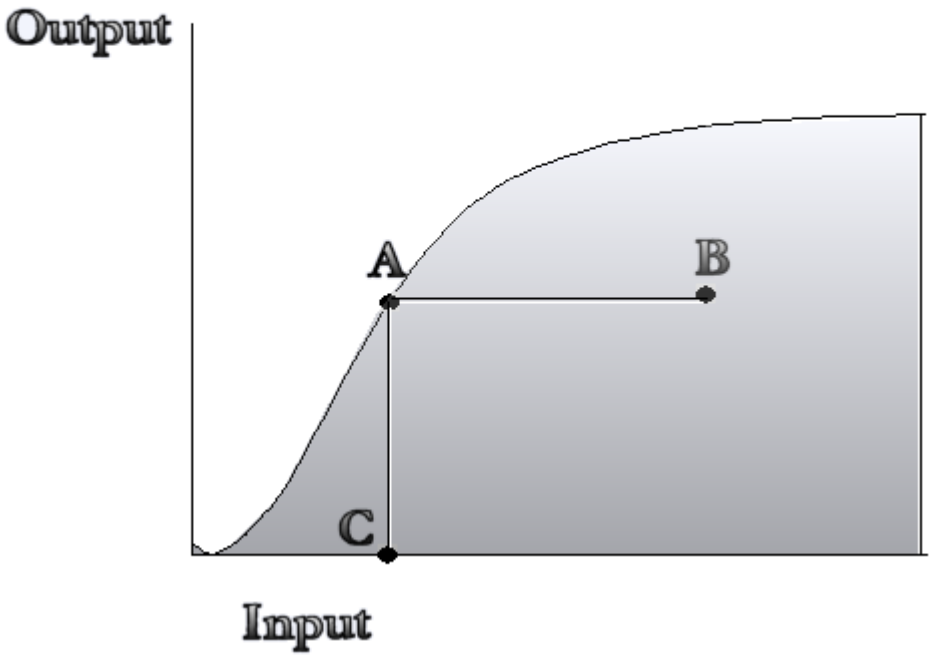
In that context, we do need to note that policy needs to be judged not just on its effectiveness, but on its cost-effectiveness. As we argued earlier in this chapter, cost is always a part of the problem definition implicitly or explicitly. Thus it is always a question of better performance at affordable cost. This is why improving efficiency, which means improving performance at any given level of cost, will often turn out to be a relevant, intermediate objective. Unfortunately, in a world of limited resources, some very important problems may not be solvable for cost reasons – as the problem of AIDS in southern Africa tragically reminds us. (Although, of course, cost is far from the *only* obstacle in that situation.)

In the process of choosing priorities, adherents of different ethical theories will tend to emphasize different concerns. Objective utilitarians will focus on the cost-effective improvement of overall health status. Subjective utilitarians' main concern will be customer satisfaction. Egalitarian liberals will want to get everyone up to some minimum level of opportunity (however that is defined) and thus will tend to focus on the health and economic status of the worst off in a society. Thus in addition to health,

financial risk protection for the poor will also matter greatly to them. Objective utilitarians will see certain community norms as obstacles to rational resource allocation – even as relativist communitarians defend them. Not everyone will agree on the relative importance of clinical versus service quality. Nor will they all put the same priority on giving patients a choice within the care process. But that is exactly why deciding on which performance problems to make priorities is a *choice*, and one that has to be made both thoughtfully and prudently.

In summary, then, problem definition – picking performance areas for priority attention for improvement – is a critical strategic decision for a health reformer. It affects how attention is focused, what interventions are tried and ultimately how the reformers themselves will be judged. It is a choice to allocate scarce political and social resources in one way and not in another. Such choices have repercussions on many levels – from the individuals and organizations involved up through to the society as a whole. And they need to be made in a self-transparent and forward-looking way. How else can a reformer establish the best possible base for the difficult work of actually getting reform to happen – which is, after all, the point of the entire process.

Figure 4.5.



Chapter 5
Political Analysis and Political Strategies

1. Introduction

Health sector reform, like all policy reform, is a profoundly political process. As we stressed in Chapter 2, astute policy developers begin political analysis early in the policy cycle: they do not delay political analysis until after the policy has been developed. Waiting to assess the political implications of a policy can affect whether it is adopted, and can thus lead to failure. The technical work of policy development and the political work of feasibility assessment need to occur *at the same time*. Political analysis and political strategies are required at each stage in the policy cycle, and are not confined to the box of 'political decision' as suggested in Figure 2.1.

Nor have advocates for health reform in developing countries adequately recognized the political challenges they confront or developed the political strategies they need. One example of limited political analysis for major health reforms is the *World Development Report* of 1993, which provided seven chapters on what countries should do to improve the "efficiency" of investments in health in poor countries, but only five paragraphs on the process of health sector reform (Reich, 1997). Furthermore, the report provided few concrete or specific ideas about how to manage what it called the "continuous and complex struggle" of health sector reform; it simply observed that "Broad reforms in the health sector are possible when there is sufficient political will and when changes to the health sector are designed and implemented by capable planners and managers" (World Bank, 1993, p. 15). *The World Health Report* of 2000, on the topic of health system performance, gave greater attention to the role of politics, including a paragraph on the importance of mobilizing stakeholders (WHO, 2000, p. 134), but subsumed politics within the broader function of "stewardship," meaning government's responsibility for social welfare and concern about trust and legitimacy (p. 119). The WHO report gave very limited attention to political analysis or political strategies, which we consider essential components of efforts to improve health system performance.

How then should health reformers go about developing and implementing political strategies to improve the chances that their plan will be adopted and implemented? Our approach emphasizes the importance of systematic and continuous political analysis. This chapter proceeds as follows. We first examine how the agenda is set for public policy in general and for health sector reform in particular. Next, we discuss how to manage this process of agenda-setting and the use of stakeholder analysis. This is a systematic way of analyzing the relevant groups and individuals inside and outside government who might influence the process of policy choice. We then present four basic political strategies for improving the chances that a policy reform will be adopted, concluding with lessons from negotiation theory about how to build a winning political coalition. Finally, we discuss some of the ethical dimensions of political strategizing, especially how a health reformer should make a personal strategic calculation.

2. Agenda-Setting for Health Reform

In the real world, what determines the health reform agenda in a given country? A naïve response might be, "Public attention focuses on areas where the performance of the health care system is unsatisfactory." But the problems defined as issues for public policy may not be those aspects that experts identify as unsatisfactory or as high priority. Consider the following example, which was mentioned in Chapter 2. In industrialized nations, women live from four to nine years longer than men, yet few nations consider this a problem for their health care systems. Indeed, in the United States, there is a major governmental effort to improve women's health, despite the fact that women in America live six years longer than men. As this example suggests, the "facts" do not just speak for themselves in setting priorities for health sector reform. More broadly, the issues selected for public policy attention do not necessarily correspond to the choices one would make based on ethical analysis or an assessment of core performance criteria (as discussed in Chapters 3 and 4).

In practice, health sector problems get defined as public issues through larger social and political processes (Reich, 1995). Issues come and go in political life as a matter of public attention, in what has been called the issue-attention cycle (Downs, 1972). Let us consider some of the important factors that shape public attention to issues in the agenda-setting process.

The mass media play an important role in shaping this cycle of issues for public debate, in both developed and developing countries.The media can transform private troubles into public issues, create awareness among the public and political elites, and shape the boundaries and symbols of public debate. The cycle of issues is shaped in part by economic incentives. Newspapers, magazines, radio, and television stations depend on revenues from circulation, either from direct sales or advertising. Readers, listeners and viewers consume the news in part as entertainment. As stories become worn out and boring over time, the media have an incentive to find new, interesting topics to attract viewers and readers. The interest of the media in specific issues can also be driven by their ownership and relationship to specific political parties, and can be limited by the state's tolerance for public criticism and basic freedoms (such as freedom of the press).

The availability of proposed solutions also affects the definition of an issue. As we noted in Chapter 2, the definition of a "problem" can be driven by the availability of a "solution." This is usually an active process, related to a committed individual or organization to promote the solution. For example, international agencies can play a key role in focusing attention on specific solutions that fit with that agency's mission, and thereby seek to set the policy agenda within a country (Reich, 1995). In the 1970s and 1980s, for instance, UNICEF made major efforts to promote GOBI (growth monitoring, oral rehydration, breast-feeding, and immunization) as its solution to high infant mortality rates. International agencies use various means to promote their policy solutions, including powerful incentives to set the agenda in the broader political system –through loans and gifts, through conditionality, and through diplomatic pressure. In these

situations, international agencies become *policy entrepreneurs* – as actors who seek to promote a particular issue and a particular solution.

Policy entrepreneurs exist within the national context as well. Consider, for example, the efforts of a group of Japanese women parliamentarians who successfully focused legislative action on child pornography (Strom, 1999), or efforts by physicians in Eastern Europe who are trying to define their own falling incomes as a health reform problem. The ability of a group to focus political and social attention on a particular aspect of the health system depends on many factors, as we discuss below, including their own resources, features of the broader environment, and political timing. The advocates of health reform, both at the national and international levels, often function as policy entrepreneurs. In this role, they need to understand the policy cycle (presented in Chapter 2) and how they can most effectively influence the policy agenda to support reform at each stage in the cycle.

A crisis often provides an opportunity to place an issue on the policy agenda. The crisis can be a natural disaster such as an earthquake, which can highlight problems in the health system that need attention, or it can be a human-caused disaster such as an economic crisis, which can focus attention on costly imported pharmaceuticals, and provide the impetus for introducing an essential-drugs policy. In some cases, policy entrepreneurs seek for years to introduce a particular policy reform, and achieve success only when the right combination of crisis and political circumstances come together, as occurred with pharmaceutical policy reform in Bangladesh in 1983 when a military dictator took power (Reich, 1994b).

Political cycles and timing also affect which issues get on a country's policy agenda. By defining specific problems as issues, different groups and individuals fight for political advantage, especially during elections, seeking recognition for their problems on the policy agenda. Once an issue has become salient, one policy suggestion tends to generate competing proposals, as competing groups claim ownership for the issue in ways that will work to the group's advantage. Once an issue is placed on the policy agenda, it pushes other issues off; a country is usually able to consider only a limited number of major social problems for policy reform at a time. Policy entrepreneurs not only must make their definition more attractive than competing definitions, but they have to make it more attractive than competing issues. For example, President George W. Bush took office in the United States with his priority issues of tax cuts and education reform, crowding other issues off the public agenda – until the attack of September 11 pushed all social issues out of public attention to focus on terrorism and the response. This example shows how political timing is critical, since the window of opportunity for policy change is often limited and subject to unanticipated events. Policy entrepreneurs need to develop an understanding of when the window of opportunity is open, how long it is likely to remain open, and how to squeeze their policy changes through the window quickly.

Changes in the policy agenda can also occur when new actors enter the political system. A new minister of health, for example, commonly seeks to define a new policy

agenda shortly after entering office. The policy agenda is intended not only to solve social problems, but also to create a political legacy for posterity and to demonstrate that the minister deserves to stay in office and to receive rewards when the term is over. How the minister defines the policy agenda will be shaped by broader political incentives and the perceived costs and benefits of specific policies, including the views held by the head of state. This process of agenda-setting by a minister of health is illustrated by the decision in Egypt to design a new policy for health insurance of school-aged children in 1983 (Nandakumar et al., 2000). In Colombia, the Minister of Health worked with Senators, to develop alliances with key legislators, in order to assure passage of his law for health reform. This example shows how a policy entrepreneur needs to consider the political circumstances of the specific policy arena, as a strategy for agenda-setting.

Policy entrepreneurs also act on the basis of their beliefs about what would make the system better – beliefs and values that are often shaped by personal experiences. Someone with a mentally ill family member may advocate better care for the mentally ill, because they have experienced poor service and the impacts on the family. A politician who has experienced a personal loss from a traffic accident may respond by pushing for legislation requiring seat belts, to enhance safety. Ideology or professional training can also influence the issues that policy entrepreneur selects. For example, public health professionals generally believe in the value of prevention, and these beliefs influence their selection and definition of problems to address.

The general culture of a society also influences the agenda-setting process, by making some topics easy to raise in public and blocking other topics as taboos. These values and beliefs tend to be specific to a particular country's culture and traditions (or a community within the country), for a particular moment in time. For example, in the United States today, it is acceptable to examine differences in health status between white and black populations, while in other countries, differences in health status among ethnic groups may be impossible to discuss in public. In some countries, it is taboo to discuss higher infant mortality rates for female children as a problem, or the existence of a free market for kidneys being sold to rich foreigners. In this it is important to remember that cultural beliefs are not static but change over time, so that topics once considered out of bounds for public debate can become defined as problems and as public issues. For example, female circumcision was considered taboo for public discussion in many countries, until after the International Conference on Population and Development in Cairo in 1995. In both intended and unintended ways, once-taboo topics can become public issues for health reform debate. Understanding the role of cultural values in the definition and selection of topics can help health reformers plan which problems they address and how, as we discuss below.

Of course, agenda-setting processes can be unpredictable. A determined, skilled, and powerful leader can make the unthinkable become thinkable, or a sudden crisis can focus attention and alter political calculations about a problem, thereby creating an unanticipated window of opportunity for policy change. Political scientist John Kingdon (1995) has argued that the best chances for successful policy change occur when three streams of events come together: 1) the objective situation, 2) the availability of a

possible solution, and 3) the flow of political events. When these three streams converge, according to Kingdon, some policy response is likely to result – although the response may not resolve the problem.

Health reformers therefore need the capacity to assess and manage the flow of political events (skills in which they are rarely trained) as well as the capacities to analyze the objective situation and design possible solutions (skills in which they usually receive training). Health reformers need to understand how agendas for public policy are set, and how they can modify the agenda-setting processes for health reform. In short, health reformers do not have to accept the problem definition generated by the agenda-setting processes in their country. Health reformers can use ethical theory and the core criteria to define the problems they think should be addressed. They then must use political processes to expand public acceptance of these normative definitions, in order to shape the policy agenda in accordance with their problem definition.

In short, health reformers must be centrally concerned with the *political feasibility* of a given policy proposal. Can the policy proposal be adopted, and can it be implemented? But the concept of feasibility is not a dichotomous, yes/no variable: feasibility involves probability that can vary from zero to one. The likelihood of getting a policy adopted depends on the situation, skill, and commitment of its advocates (and its opponents). What political resources do they have at their disposal and how much are they prepared to commit to this particular policy battle? How clever and effective are they at such diverse tasks as assessing the political consequences of technical proposals, making emotional public appeals, and negotiating private deals with key political actors? How do you persuade individuals and organizations to commit their limited political resources and capital for this particular reform effort? Leadership here is critical. Talented political leaders, like talented generals, can win battles and campaigns that would overwhelm those less able or less energetic. So, asking if a policy is *feasible* is, in part, asking a question about the advocates of reform, especially their creativity, commitment, and skills – and about their opponents. Passing health care reform often means overcoming or deflecting powerful interests who will vigorously defend their positions.

Constructing health reform so that it is politically feasible thus requires political skills, political analysis, and political strategies – rather than some vague notion of "political will," as the World Bank called for in its 1993 *World Development Report*. Players involved in health reform make their own calculations of the likely political costs and benefits of reform, and take positions and spend resources accordingly. Advocates of health reform need to understand these calculations from each player's perspective, and then construct strategies that will influence these calculations and tip the probabilities in favor of reform. This process requires the collection and analysis of political data, including subjective judgments about how different players are likely to respond to different reform proposals. It therefore requires a good understanding of the politics of health sector reform.

3. Politics of Health Sector Reform

We believe that four factors are of particular importance and can provide practical guidance in efforts to manage the politics of health sector reform – especially how political processes affect the origins of health reform as an issue, the content of reform policies, and the implementation of adopted programs. The four factors are:

- *Players*: the set of individuals and groups who are involved in the reform process, or who might enter the debate over the policy's fate;

- *Power*: the relative power of each player in the political game (based on the political resources available to each player);

- *Position*: the position taken by each player, including whether the player supports or opposes the policy, and the intensity of commitment toward the policy for each player (i.e., the proportion of resources that the player is willing to expend on the policy);

- *Perception*: the public perception of the policy, including the definition of the problem and the solution, and the material and symbolic consequences for particular players.

These four factors – players, power, position, and perception – can be influenced through the political strategies adopted by health reformers. A central purpose of this chapter is to provide basic guidance on how to design those political strategies.

Of course, the political feasibility of a proposed policy is affected by other factors beyond these four. Changing the content of a policy can affect the distribution of political costs and benefits among players, and thereby alter the calculus of political feasibility. In addition, the strategies adopted by key players can alter the behavior of other players: one player can make decisions that alter the power and position of other players; sometimes those decisions change the rules of the game itself. At the same time, some factors are difficult or impossible for health reformers to modify or manipulate, including deep-rooted cultural beliefs, the incentives presented by international agencies, and the occurrence of natural or economic disasters.

What would be an appropriate metaphor for policymaking and for thinking about political feasibility? First, policymaking is not like a game of checkers, where all players are equal, and can only move in predetermined ways. It is more like a game of chess, with different moves for different players, and where pawns can become kings if they reach the other side. But there is a major difference: in policymaking, the rules of the game are not stable, and can change (sometimes imperceptibly and sometimes explicitly) as the game proceeds. A policy's feasibility can be determined by the rules and who controls them: no policy is assured of continuity.

In addition, the politics of health sector reform has certain *systematic characteristics* that make it a difficult process – perhaps even more so than other kinds of policy reform:

- *Technical complexity/difficulty*: fixing the health sector is not easy. Many parts and pieces are interrelated, and many consequences (both intended and unintended) occur. Designing a comprehensive health reform is a complex technical process, as health reformers turn the five control knobs in various directions. Reformers often seek to improve many parts of the system at the same time, making both the details and the overall impact of the program difficult for nonexperts to grasp. These technical challenges create political challenges through the impacts on players and the potential for public confusion.

- *Concentrated costs on well-organized groups*: health sector reform efforts commonly place concentrated new costs on well-mobilized, powerful groups; for example, on physicians (often well-organized in a national medical association), or on the pharmaceutical industry (often well-organized in an industry association). This problem of concentrated costs can create significant political obstacles to reform, if the high-power groups become mobilized to oppose the reform, to protect their interests.

- *Dispersed benefits on non-organized groups*: health sector reform often seeks to make new benefits available to previously disadvantaged groups; for example, the poor, or rural residents. Such groups often are not well organized or politically well connected. In addition, these changes may only result in modest benefits for each individual. Dispersed benefits among low-power groups make it more difficult to mobilize significant political support for reform.

The combination of concentrated costs on well-organized groups and dispersed benefits on non-organized groups constitutes what Mancur Olson (1965) called a collective action dilemma. Overcoming the politics of this collective action dilemma is a major challenge for health reformers. Together, these three characteristics of health sector reform create considerable political challenges; at the same time they also allow common approaches to the politics of health reform, especially in the conduct of stakeholder analysis and the design of political strategies.[17]

4. Political Analysis

A basic building block in designing political strategies for reform is to perform a stakeholder analysis (Reich, 1996). This consists of three stages. First, identify the relevant groups and individuals. Second, assess their political resources and their roles in the political structure, to determine their relative power for the policy question at hand. Third, evaluate their current position on the proposed policy (including the intensity of

[17] For a computer software program that helps with both stakeholder analysis and political strategies, see *PolicyMaker*, by Reich and Cooper (1996).

their commitment) and their underlying interests. This analysis can be presented visually, as shown in Figure 5.1 for the case of health reform in the Dominican Republic (Glassman et al., 1999). Once the stakeholder analysis is completed, with a clear understanding of the relevant groups and their power and positions, a health reformer can proceed to the next stage – designing political strategies to enhance the chances that the reform will be adopted.

Obtaining reliable information for the stakeholder analysis – about the policy position and power of different players – can be quite difficult. Sophisticated political players may purposefully mislead others about their position, in order to gain advantages in the negotiations over critical policy issues – and sometimes they play on both sides of the debate. Similarly, stakeholders have an interest in masking their power or in exaggerating their power, to gain added influence in negotiations. Even if you can assess a player's power with some accuracy, it can be difficult to know how much they are willing to use on a given issue. The public positions of key players can be determined, through public statements, but those often do not represent the actual positions that players will take when pressed for a decision. The most important task for analysts is to determine the core position and power of stakeholders – what they are not willing to give up, their final offer for a compromise, and the amount of resources they are willing to mobilize to assure their final position. Finding out this information requires multiple sources and careful judgment about specific players – questions usefully considered by a team of analysts with differing perspectives.

Stakeholder analysis combines two distinct modes of analysis within political science. One is interest group analysis. This consists of understanding those social groups that are seeking to press government in a particular direction, including private business and non-governmental organizations. The second mode of analysis examines bureaucratic politics, and is focused on the competition among agencies and individuals within government. The stakeholders relevant to health sector reform include both kinds of players: those outside government and those inside government. For health sector reform in developing countries, one also needs to consider the activities of international agencies, including the World Bank and the World Health Organization.

Health sector reform typically involves the following kinds of interest groups:

- *Producer groups*: doctors, dentists, nurses, pharmacists, other health sector employees and their unions, domestic and international pharmaceutical companies, and equipment manufacturers;

- *Consumer groups*: disease-based organizations, local and regional consumer groups, women's organizations, unions representing insured workers, retirees, and military groups;

- *Economic groups*: businesses with health insurance schemes, industries affected by health policies (e.g., tobacco farmers, drug sellers), and workers who gain or lose jobs;

- *Ideological groups*: political parties, reform organizations, single-issue advocates (e.g., environmentalists, anti-abortion activists);

- *Health development groups*: multilateral development banks, bilateral aid agencies, international health organizations, non-governmental development organizations.

This list of interest groups is illustrative. Health reformers need to adapt the list to the specific political circumstances for the policy at hand, and focus attention on the most critical groups and individuals. Deciding on the list of interest group players for analysis requires judgment about the groups most likely to be mobilized or that could be mobilized to affect the balance of power in the policy debate.

Similarly, when examining bureaucratic groups, a political analyst must decide which actors to consider and where to draw the lines. Is the ministry of health a single player, or are different parts, like the minister and his immediate staff, best thought of as separate groups? There is no one right answer to such questions. Carrying out a good political analysis is an art as well as a science. The goal is to identify the important players who can act independently and can control significant political resources. But considering too many separate players makes the analysis unwieldy and frustrating.

In addition to the Ministry of Health (MOH), the following government actors often play important roles in health sector reform debates and decisions:

- *Ministry of Finance*: especially if the reform involves changes in the financing of health services or changes in the MOH's budget;

- *Social Security Institute:* especially if the reform proposal involves changes in the provision of health care and a reorganization of government health facilities;

- *Economic and/or Planning Ministries*: especially if the reform proposal involves calculations about overall economic growth or decisions about debt forgiveness;

- *Local and/or Regional governments*: especially if the reform proposal involves decentralization;

- *Ministry of Education*: for school health policies or for policies that affect medical schools;

- *Ministry of Agriculture, Trade, and Industry*: for policies on alcohol, tobacco, pharmaceuticals, or medical equipment.

In compiling the list of stakeholders, the analyst needs to consider the implications of the proposed policy for each player: Who cares about the policy? Who is likely to act – or could be convinced to act? Who has the potential to influence the

outcome? Who is likely to be affected by the reform's consequences, in positive and negative ways?

The analyst also must consider critical political actors: individuals or groups who have special power over the policy in question. In some cases, a powerful individual may have veto power over the adoption of a policy reform; in many cases, a particular group can determine whether a policy is implemented at all. Leaders deserve particular attention. The prime minister or president, specific senators or governors, leaders of important parliamentary factions and political parties: all may well matter. Media outlets also must be examined, especially those organizations strongly identified with particular points of view, parties, or factions. The editors or managers of media organizations often have their own political agendas, and they need to be considered in the political analysis of health sector reform.

4.1. Sources of Power and Influence

Once you have identified a list of players (the groups and individuals who are relevant to the policy proposal), the next step is to estimate the power and influence of each player. This requires an assessment of the following for each player.

- The player's political resources and place in the political system, which determine the potential capacity to influence policy decisions;

- The player's interests, position, and commitment, which will influence how the player's resources – and how much of those resources – will be used in the policy debate.

Political resources come in two forms: tangible and intangible (see Figure 5.2). Tangible resources include money, organization, people, votes, equipment and offices, all of which affect a group's ability to influence the policy process. Money can provide political contributions and purchase other resources such as expertise, media access, and organizational support. People can provide workers to distribute literature, participate in rallies and demonstrations, and help with lobbying government officials on particular policy issues. Votes can matter, depending on whether a decision is to be made through an electoral process, in a legislature or committee or a popular election. The underlying level of organization also affects a group's influence. Groups with existing infrastructure (offices, staff, electricity, telephone services, fax machines, and computers) have a substantial advantage in the policy process. These groups have already paid the transaction costs of organizing themselves and can utilize that investment to influence the issue at hand. A new group, on the other hand, first has to pay the costs of becoming organized: it has to recruit members, arrange a staff, develop materials, and construct systems. These costs reinforce the anti-change bias and inertia of political systems, because new groups (which could arise with a new policy) are not yet organized to express and defend their interests.

The second set of political resources is intangible. These resources include information on the policy and the problem, as well as relevant substantive expertise, which allow the group to develop a position. Two important intangible resources are the group's visibility and legitimacy. Groups whose members have access to key decision makers, whether from personal contacts or past favors, do better than those without good connections to powerholders. Health reformers also require the political skills to manage the lobbying process. Knowing where to go, whom to talk to, and how to get the press interested – all enhance a group's policy impact.

This analysis of political resources helps explain why particular groups play the roles they do in shaping health sector reform. Physicians tend to be influential because they are usually well organized, wealthy, skillful, expert, and seen as legitimate by other players; moreover, every top decision maker has a personal physician. Consequently, the medical association is often well placed to put its issues on the political agenda for consideration and to shape policies in the health sector. On the other hand, international pharmaceutical companies may have large money resources and good access, but they tend to have limited grassroots labor and limited social legitimacy. Groups that are not well organized (such as the rural poor) confront major obstacles before they can begin to articulate their interests in an effective way. In short, political power and political resources are not equally distributed in society – and that distribution affects how the policy process works for health reform. The distribution of power and resources shapes the collective action dilemma mentioned above, which creates a fundamental political obstacle to reform.

The value of different kinds of political resources depends partly on the nature of the political system. The degree of democracy in a country, as expressed in freedom of speech and freedom of association, can constrain group activities in fundamental ways. These factors shape the distribution of power in society and the kinds of participation allowed in the policy process. In many developing countries, interest groups are not well organized, and the ruling elite constrains the formation and activities of interest groups through coercion or legislation. Still, for health policy, the physicians association, trade unions, and the pharmaceutical manufacturers association often exert significant influence on policies that affect their interests. For these reasons, it is important to consider how the existing health system distributes both economic and political benefits to certain interest groups, and restricts benefits to other groups (Reich, 1994a).

As part of their political analysis of the policy debate, health reformers can conduct an inventory of political resources held by different players. Differentiating tangible from intangible resources is useful because players build their strategies based on their resources. This means that players weak in material resources (for example, with limited money and equipment) will often rely on political strategies that depend on intangible resources (for example, protest actions or challenges to legitimacy). You should know your own political resources, and build your strategies accordingly; you should also assess your opponents' resources, in order to predict their likely strategies. Even though you lack material resources, you can still design political strategies that can give you substantial leverage in a policy debate, by wisely using symbols that can

connect to broad social values. On the other hand, powerful political actors can use symbolic strategies to prevent the serious consideration of a new policy issue, by linking their position on a policy to deeply held cultural worldviews in ways that control the policy agenda (Cobb and Ross, 1997).

Elements of political structure also influence the leverage of different social groups. For example, in the U.S. political system, media-based electoral campaigns are essential, which increases the influence of those with money, as money is vital for television advertising in elections and policy debates. In countries without competitive elections, or where the state controls radio and television broadcasting, money is less important for purchasing air time to promote a particular policy position. However, in a corrupt political system, money often is important for purchasing access to decision makers or for purchasing a specific policy decision. In parliamentary systems, proportional representation can force political leaders to give more attention to minorities than occurs in legislatures with winner-take-all single member constituencies. On the other hand, single-member constituency systems give an advantage to geographically concentrated groups – who might control swing seats. But those systems also create obstacles to the influence of dispersed groups of committed believers (who have more impact under proportional representation).

A group's role in the existing pattern of political competition can affect the group's influence. An important distinction is between *swing* and *base* groups. A swing group is one that might vote for either party in an election, or a legislative bloc that might be for or against the government. A base group, in contrast, is a firm member of a party's core support.

For many interest groups, the strategic question is how to convince the ruling coalition to pay attention to the group's views. In some cases, an interest group may seek to keep its loyalty uncertain, in order to induce competing parties to bid for the group's support. For example, powerful health care interests in the U.S., seeking to send this message, contribute heavily to both political parties. Political leaders, on the other hand, seek to expand their political support by relying on the loyal support of their base groups, while trying to attract the conditional loyalty of swing groups. Reaching out too far to attract support from a swing group, however, can threaten the ongoing support of a base group, and persuade a politician to pull back.

Interest groups thus use their role in the political system as a resource for seeking concessions from politicians in either the ruling coalition or the opposition at particular times. During elections, groups may be able to obtain concessions or promises on policies that may be difficult to elicit under normal circumstances. Elections can therefore provide opportunities for relatively powerless groups to expand their influence on health policy issues, such as the distribution of health facilities or access to health services. Furthermore, an ongoing political struggle within a country, or an economic crisis, can change the distribution of power and present opportunities for interest groups and bureaucratic actors to shape policy decisions in the health sector.

4.2 Position and Commitment

Knowing a group's political resources does not enable one to predict whether or how those resources will be used. Stakeholder analysis therefore needs to assess how each group views the policy issue, and what its position is.

A first step is to assess the group's interests. These are not always clear even to the group itself, nor does a group necessarily pursue its own interests. But in many cases, interests do shape policy positions. For example, knowing that physicians' incomes will decrease under a proposed payment plan strongly suggests that the medical association is likely to oppose the plan. Economic consequences of a policy often determine political positions on that policy.

It is also important to examine the positions that the players have taken publicly. Do the positions correspond with their interests as we understand them? If not, how can the discrepancy be explained? Is it because some groups are confused, because they are proceeding tactically, or because they know something that we do not? Differences can also arise between public positions and private positions, between what a group says it will do and what a group will actually do. In addition, ideology may not be a good predictor of position, since political action often involves both interests and willingness to bargain and compromise as well as values. For example, in Costa Rica the trade unions and the communists were willing to support some efforts to privatize social security, in order to introduce competition and make the bureaucracy more interested in efficiency and quality.

As part of this analysis, we also need to describe the intensity of each group's current position. How much of their available resources is the group willing to use to promote its position? For example, how much time is the group's leader spending on the issue? Is the group expending a significant portion of its financial resources on the issue? Is this an issue that the group is prepared to fight for to the death, or is the commitment tepid, less resource intensive, more of "just going through the motions"?

Developing effective political strategies requires an understanding of why the key groups are taking their positions. Here are two hypothetical examples. Suppose the ministry of finance is opposing an increase in the tobacco tax that is intended to support rural health care. Is that a result of a general bureaucratic reflex, or because the minister has a different, specific use for the money in mind? If the latter, then how important is that alternative to the minister? A minister acting out of a general desire to protect his turf poses a different political problem than someone strongly committed to an alternative program. Or suppose the medical association is vigorously opposing the new fee schedule in the national health insurance plan. Is that because most of its members genuinely understand and oppose the new system, or is a small group of successful specialists driving that position? If the latter, then the association's leaders might be willing to make a symbolic protest to satisfy an internal constituency, and they might accept a deal if most association members could be protected from adverse effects. On the other hand, if the association's president is driving the opposition to the policy, it may be difficult to

change the official position as long as that individual remains in charge, or unless the individual has a specific request that can be met.

The intensity of group mobilization thus depends in part on how the policy is expected to affect the group's interests and how those likely consequences are perceived. Decisions that are perceived as affecting members of a well-organized group (with relatively few members) often result in high mobilization. For example, a government policy to restrict imports of high-cost medicines through an essential-drugs policy would probably result in strong opposition from multinational pharmaceutical companies and their industry association. On the other hand, decisions that are perceived as having marginal impacts on the members of poorly organized groups may not result in much mobilization. For example, a policy to increase user fees for patients at rural clinics may not result in protests by patients from rural areas.

The positions of the major political parties can be critical to health sector reform. Because parties are coalitions, bargaining is a core part of the process of developing programs and positions. If a party has already embraced health sector reform as a priority issue for its agenda, then much of the work for reform advocates has already been accomplished. Using parties to help manage the political issue of health reform is always an option to be considered in the development of political strategies, the topic that we discuss next.

5. Political Strategies for Reform

How can a health reformer develop political strategies for forging a winning coalition? We noted above that the political feasibility of health sector reform (and policy reform more generally) is shaped by the *position* of the players, the *power* of each player, the number of *players* mobilized, and the *perception* of the problem and the solution. Reform advocates need political strategies to manage each of these factors, in order to increase the probability that their reform plan will be adopted and implemented. Below we consider four sets of political strategies for reform.

5.1. Strategy #1: Bargain to Change the Position of Players

The first set of political strategies involves bargaining within the existing distribution of power, to change the position of players. This can involve deals, promises, trades, and threats.

One bargaining approach is to offer to change a specific element of the policy's substance as an incentive to move specific actors more in favor, to switch sides, or to be less opposed. Perhaps the minister of finance will agree not to oppose the proposal for higher cigarette taxes for rural health care, if only part of the revenue is used for health care and the rest is left for other government purposes (possibly including the minister's personal projects). Or, in another example, suppose that a country proposes changing from per-day to per-admission payments for hospitals. Perhaps the owners of private

hospitals will not oppose the new payment system if, for the first three years, there is a hold-harmless rule that guarantees that no institution will lose revenue for that time period.

A second kind of bargaining is inter-issue exchanges, in which an opponent agrees to cooperate on one issue in return for something else they value on a different issue. Such "logrolling" is normal in many political contexts. If the medical association accepts the new fee schedule, perhaps the ministry will agree to accept new limits on the number of doctors who can practice in crowded (i.e., desirable) areas. Or suppose that provincial governments ask the central government to take on responsibility for financially troubled provincial hospitals – in return, they might offer to accept a different formula for dividing the national tax revenue so that more goes to the central government. While the trades in these examples are on related issues, that does not have to be the case: maybe the unions will accept an increase in payroll taxes to support the health insurance scheme, if the government promises some other efforts to increase overall employment.

Bargains can involve threats as well as promises: doctors may threaten a strike if their salaries are not raised; consumers can threaten demonstrations if their health benefits are reduced; pharmaceutical companies can threaten to close their plants if the country implements a price-control scheme. Sometimes, the threat is to stop cooperating. For instance, small business owners might say they will refuse to report earnings and will stop paying taxes on those earnings for the new health insurance scheme, unless their overall contribution ceiling is lowered.

Constructing an effective bargain to change the positions of key players requires intelligence, skill, and trial-and-error experimentation. Skills in negotiation and conflict resolution are particularly important for these strategies. Finding out what it takes to change a player's position may require an actual offer or threat. The bargain could involve modifying the policy's content; in that case, the health reformer should determine how the change influences the policy's expected effectiveness. A compromise that produces a supportive coalition but undermines the reform's overall goals could create a reform that is politically feasible but technically ineffective or even counterproductive. For someone who cares about consequences (such as our core performance criteria), that kind of compromise would not be a good deal.

5.2. Strategy #2: Distribute Power Resources to Strengthen Friends and Weaken Enemies

The second set of political strategies is designed to change the distribution of power among key players. Since a group's impact on the policy process depends partly on its resources, reform advocates can adopt strategies to enhance the tangible and intangible political resources of supporters and decrease those resources of opponents. Here are some examples:

- Give or lend money, staff, or facilities to groups that support the reform;

- Provide information and education to supporters to increase their expertise;

- Give allies expanded access to lobby key decision makers;

- Provide allies with media time and attention to enhance their legitimacy; focus attention on their expertise, impartiality, national loyalty, and other positive social values.

A tough political strategist can also do the same – in reverse – to opponents:

- Decrease opponents' resources by inducing people not to contribute and not to work for them; do this by attacking the group's legitimacy, honesty, or motives;

- Refuse to cooperate with opponents, for example, by not providing information; keep them uninformed and out of the loop;

- Reduce opponents' access to key decision makers;

- Urge the media not to cover opponents; characterize them as inexpert, self-serving, disloyal, etc.

These power strategies may not be effective, or ethical, in all cases. The list above illustrates some examples of the kinds of power resources that a strategist can consider, rather than an inclusive list of what should always be done. Deciding on which power strategies to select requires an assessment of the players involved in health reform (both for and against), the kinds of power resources available to health reform advocates, and judgments about the likely impacts of different power strategies on specific players.

5.3. Strategy #3: Change the Number of Players, by Creating New Friends and Discouraging Foes

Reform advocates can also consider political strategies that try to change the set of friends and foes. These strategies seek to mobilize players who are not yet organized and demobilize players who are already organized. It means changing the number of mobilized players, as supporters and opponents, by recruiting political actors to the health reform cause, and away from the side of the opponents.

Mobilizing groups requires convincing people that they should pay the non-trivial costs of getting involved in an issue they have so far ignored, or the substantial costs of organizing a new group. Sometimes, mobilizing a new group may simply require bringing the issue to the group's attention – once the group knows what is going on, it may decide to take a position. If the nation's medical students have not begun to protest a proposed tuition increase, it may be because they have not yet heard of the proposal. Informing the student group (or its leaders) about the proposal and its consequences may be sufficient to mobilize students as a support group.

Mobilizing the unorganized is more complicated, as the group may not understand their interest in the issue, until the policy and its consequences are explained to them. For instance, the unions may think that a new competitive private health insurance scheme is acceptable, until someone explains that it will significantly increase administrative costs and thereby increase the premiums paid by their members. In some cases, competing factions within a group may have different views on the group's interests and values, and on the policy's consequences.

Nor is demobilizing a group that has taken a public position easy, for reasons of face and interests. In some cases, a group can leave the policy debate through a discussion of its key interests in which there may be a chance to explain to the group that its stake in the issue is different from what they thought it was, and that other issues should be of more concern to them.

Health reformers should also consider the potential to mobilize or discourage political leaders. To them, the argument is often a mixture of both self- and public-interest considerations. Show them how supporting you (or opposing you) will work to their political advantage (or disability), and why this is a good (or a bad) issue for them to get involved in.

Merit-based arguments about the public interest can sometimes be as powerful as arguments based on political interests. Here is a generic example: "Senator, I know your constituents want you to do this, but they are being short-sighted, for it really is a bad thing for the nation for the following reasons ..... So, perhaps you need to make a speech or two against this subject for the sake of the people in your district, but at the same time you don't need to actively work against us on this policy."

On the negative side, it is important to consider how to divide or undermine coalitions that are opposing you. Suppose that the medical association has decided to oppose a new insurance scheme because it will seek to contain reimbursement for high-cost procedures, which would negatively affect the income for physicians. It may be possible to convince doctors who provide primary care to switch sides and support the plan, and thereby divide the medical association, if they can be persuaded to see their interests in a different light.

A final way to change the number of mobilized players is to change the arena of decision-making. Moving the policy decision from the executive branch to the legislature, for example, can allow advocates to mobilize new groups of politicians, or to bypass political obstacles to reform. This occurred in the example noted above for Colombia, when the Minister of Health moved the debate to the legislature and developed alliances with key Senators, in order to assure passage of his reform law. Shifting the policy arena, however, can also have unanticipated consequences. For example, when 39 pharmaceutical companies challenged a South African medicines law in court, seeking to protect their patent rights for AIDS medicines, activists mobilized non-governmental groups around the world in opposition to the court case. The companies eventually

retracted the case. In this instance, the companies miscalculated the policy battle not in terms of legal argument but in terms of political mobilization and public opinion.

Political strategies can thus be used to alter the number of mobilized players and thereby influence the political feasibility of health reform. New players can enter the policy debate and take controlling positions, and current players can leave, become inactive, or wait on the sidelines. Issue advocates who want the political system to decide in favor of health sector reform need to consider all such options, and decide on strategies that change the balance of mobilized players.

5.4. Strategy #4: Change the Perception of the Problem and the Solution

So far, we have discussed political strategy with a focus on stakeholders: the players, their power, and their positions. In many contexts, this approach may be the most useful and pragmatic. For example, in a one-party political system, most of the relevant political competition will be among well-organized interest groups and established bureaucratic actors, occurring within limited processes of consultation. In such situations, direct appeals to the public are unlikely to influence political decision makers. But in other political systems, particularly those that are open and competitive, a public appeal to change the perception of an issue can be an effective political strategy for influencing bureaucratic and political leaders, as well as mass audiences. Even in relatively closed political systems, reframing an issue can change the positions and power of key players and thereby affect the political feasibility of reform.

Political strategies directed at perceptions seek to change how people think and talk about health sector reform, how the issue is defined and framed, and which values are at stake. The perception of an issue also affects how it is connected (or not) to important national symbols or values: Is this reform going to advance the nation's identity in some fundamental way?

Perception strategies relate to how the human mind works. Human beings often have trouble grasping complex fact patterns and seek ways to make sense of a confusing reality. This is especially true in situations where reality is complex, outcomes are uncertain, and conflicting goals are involved – all of which occur for health sector reform. In such cases, health reformers need to manage public perceptions, because these change how problems are defined and which solutions are acceptable. At the center of the political debate is a contest over the meaning of images and words, the symbols for health reform (Edelman, 1977). How is the problem characterized, how are the choices described, and how is the issue framed?

Consider the following example. A country is proposing to introduce private health insurance, which would allow citizens to take their contributions out of the monopoly national insurance scheme. The opponents characterize this proposal as a disastrous decline in solidarity and a loss of protection, since coverage under the insurance scheme would be decreased for the first time in a generation. On the other side, supporters characterize the change as an increase in freedom for individual subscribers

and a challenge to the ineffective state bureaucracy. Depending on which problem definition becomes widely accepted, the public and key players will view the reform effort in starkly different terms. In short, reframing an issue can change political circumstances in major ways, with significant consequences for the mobilized players, their power, and their positions.

How does one alter public perception of an issue? As noted above, the contest over perception involves conflict over who will control the meaning of key symbols of health reform. An advocate can argue that an opponent's problem definition is based on poor data, that the problem is unimportant compared to other national priorities, that the problem is not fixable at a feasible cost, or that the problem should be a matter of private not public responsibility. Identifying some citizens as undeserving victims of evil forces can also be an effective way of reframing the public debate and altering the political balance. Another technique is to invoke important national symbols that connect to broadly accepted social values, as political mobilization is about emotion as well as data, and values can be used to oppose as well as promote health sector reform.

Strategies based on core social values can also be used to defeat health reform, as illustrated by the debate over the Clinton health reform in the United States in the early 1990s. Opponents to the Clinton reform developed a series of television commercials around Harry and Louise, a fictional middle-class married couple, who presented the Clinton plan as threatening their lives. Sponsored by the health insurance industry, these commercials raised deep fears that the Clinton plan would limit the freedom of choice for existing health insurance and would produce a "government-run" health system (Johnson and Broder, 1996, p. 16). This campaign connected to deeply felt social values in the American middle-class, including anti-government sentiments, and fears of an eroding standard of living.

In contrast, the proponents of the Clinton health reform failed in the arena of public perceptions. Their reform proposal was so complicated that it defied simple explanation. For instance, they created new institutions for managed competition that were first called "health insurance purchasing cooperatives," or HIPCs – an idea that few people could understand, pronounce, or remember. They were then renamed "health alliances," which sounded more friendly but still remained obscure to most people. Furthermore, reform proponents failed to find effective symbols to explain how the Clinton plan would work, what the plan would mainly do, or how it would connect to core social values. As one key technocrat later reflected, "Many people couldn't understand what we were proposing. There were too many parts, too many new ideas, even for many policy experts to keep straight" (Starr, 1995, p. 25).

Health reformers, trained in technical skills, are often uneasy at designing political strategies for shaping public perceptions, and are often not prepared to implement them. But these perception strategies are essential for changing the health sector, as part of the overall political strategies for promoting reform. And if the supporters of reform neglect the domain of perception strategies, they are likely to face

those strategies wielded by their opponents – as demonstrated by the experience of the failed Clinton reform.

6. Negotiation and Political Strategies

Each of the four kinds of political strategies requires a process of negotiation. In recent years, systematic research on such processes has identified methods for conducting negotiations successfully. Here we review those lessons about negotiation that are relevant to political strategies for health sector reform.

The first point is that *negotiations are often successful when the issue is constructed in win-win terms* rather than win-lose terms. A key challenge in negotiation is to seek *value-creating* solutions, in which enough participants are willing to support the deal, producing win-win outcomes. The contrasting approach of "I win, you lose" represents a *value-dividing* negotiation, which sharpens differences and makes agreement more difficult. In our discussion of bargaining strategies above, we gave several examples of deals where advocates could come up with a better situation than the status quo, and where potential opponents gained. For example, if the reform plan will create new revenue sources, then potential gains can be divided to enhance the potential of a win-win situation. In Egypt, health insurance for school children was financed largely by a new tax on cigarettes, which made the proposal acceptable to key politicians in parliament who were concerned about equity impacts of premium payments on poor rural families with many children (Nandakumar et al., 2000). Of course, finding new revenue sources is not always possible, and a winning coalition does not have to include everyone. On the other hand, win-lose negotiations can transform potential supporters into active opponents, as illustrated by the unsuccessful Clinton health reform process.

A second lesson about negotiation, in looking for a win-win agreement, is to *respond to the real interests of other parties, not to the positions they happen to take*. After all, a position may well be a stance designed to take advantage in the negotiations. The hospitals in a country may say they will never accept budget responsibility, but their opposition may depend on their lack of authority to manage their own costs – for example, with regard to hiring and firing. If a negotiator knows that expanded autonomy is in the interests of hospitals, then their stated position can be reinterpreted. Offering them what is really in their interest may bring them into the support coalition.

A third lesson is that *some value-dividing conflicts may be inevitable*. For example, no matter how one reframes the issue, physicians may want to be paid more, while the ministry wants to pay them less. In such situations, one approach is to substitute *principle-based* negotiation for specific substantive confrontation, in which parties try to agree in advance to a set of principles that they will use to settle disagreements before they deal with the problems in detail. For example, they can agree on how much of the national budget should go for health or what the payroll tax should be, and then agree on what fraction should go for salaries, before they talk about specific items in the fee schedule. This approach can soften the sharp edge of confrontation and opposition, since

once an agreement on general principles has been reached, some of the details may follow accordingly.

A fourth guiding principle is for negotiators to *cast the emotional tone and energy in a positive instead of a negative direction* – or at least seek to establish a neutral and professional environment. Negotiation is a human process, filled with emotion as well as logic. Negotiators, therefore, need to manage the emotional dimensions as well as the logical content, for success.

The tone of negotiations is important because reform rarely is a one-time event; reform usually is an ongoing process, a series of connected conversations, as policies are designed, adopted, implemented and modified over time. Moves in one round of bargaining have consequences for subsequent rounds, and can be understood as moves in a long series of games. In this series of negotiations, successful agreements make possible future successful agreements, and the converse is also true. A critical dimension is trust among the players. For example, different industries in different countries show large variations in the climate of labor relations. Some employers and unions are always at each other's throats, while others resolve problems in a cooperative manner: through repeated negotiations, they have become engaged in mutually reinforcing cycles of trust or distrust over time. The same can happen between nations. When political commentators discuss momentum in the Middle East peace process, they are talking about rising or falling relationships of trust. Similarly, health sector reformers need to consider the role of trust in making it easier or harder to achieve their goals.

Conducting negotiations also requires an understanding of yourself and others. You cannot make proposals focused on other person's interests, unless you know what those interests are; you cannot know what deals you can offer without undermining your own core goals, unless you first understand your own objectives; you cannot identify which groups might constitute a winning coalition, unless you have made a systematic analysis of all possible supporters and opponents, and their positions and power. You need to understand that initial offers on both sides often represent maximal positions. And you should clearly define your own bottom line, the Best Alternative To No Agreement (BATNA), the last offer when you are willing to say yes and beyond that when you are prepared to walk away from the negotiations.

The analysis for negotiation requires a serious investment of time and energy. As we said above, health sector reform is a profoundly political process. The choice is not whether to negotiate, but rather how well prepared a health reformer is to negotiate in terms of intelligence, skills, and stamina.

7. Political Strategies and Ethics

The politics of health reform should not be approached solely as an instrumental or Machiavellian problem; health reformers need to consider the ethical dimensions of their political strategies, as well as how to get the policies they want. Values shape the substantive content of health reform, as we have emphasized in prior chapters. But values also shape the processes for promoting the adoption of the health reform plan. How then can a health reformer assess the ethics of the political process for reform?

One approach is to evaluate the anticipated gains from reform, compared to the expected costs of achieving the reform – a kind of utilitarian calculation. After developing a health reform plan and designing a set of political strategies, an advocate could ask, "How likely is it that the effort will succeed? And are the gains of achieving this reform worth the costs in time, effort, money, emotion, loyalty, and conflict that such an effort would entail?" A realistic analysis of the latter question would recognize that human beings have different kinds of goals, and that these goals will not always coincide. Most people have personal goals for career advancement, financial security, power, and influence; we also have broad goals for making our society a better place (as discussed in Chapter 3); and between these, we often have goals for intermediate organizations and groups – we may feel loyalty to the ministry, to the medical profession, or to the university we attended.

The easy decisions are when the goals all line up. If the nation, the ministry, the medical profession, and I, personally, will all benefit from a policy, then it is easy to pursue reform aggressively. But what happens when the goals do not line up? Suppose I believe that health sector reform will help the country, but could destroy my career (or my family) if I commit myself to fully supporting it? Or suppose I believe that the nation would be better off if many of the functions of my current agency were privatized or taken over by others? This would be a terrible blow not only to me, but also to many friends and colleagues. Making these decisions involves difficult soul-searching, just as evaluating the costs and benefits is fraught with uncertainties and intangibles (as we noted in our discussion of utilitarianism in Chapter 3).

Another approach to the ethics of reform processes emphasizes the role of transparency and democratic deliberation as important considerations for technical analysis and technical experts. Here we are asserting our own normative assumptions about how the policy process should be conducted, and suggesting the implications for health reform.

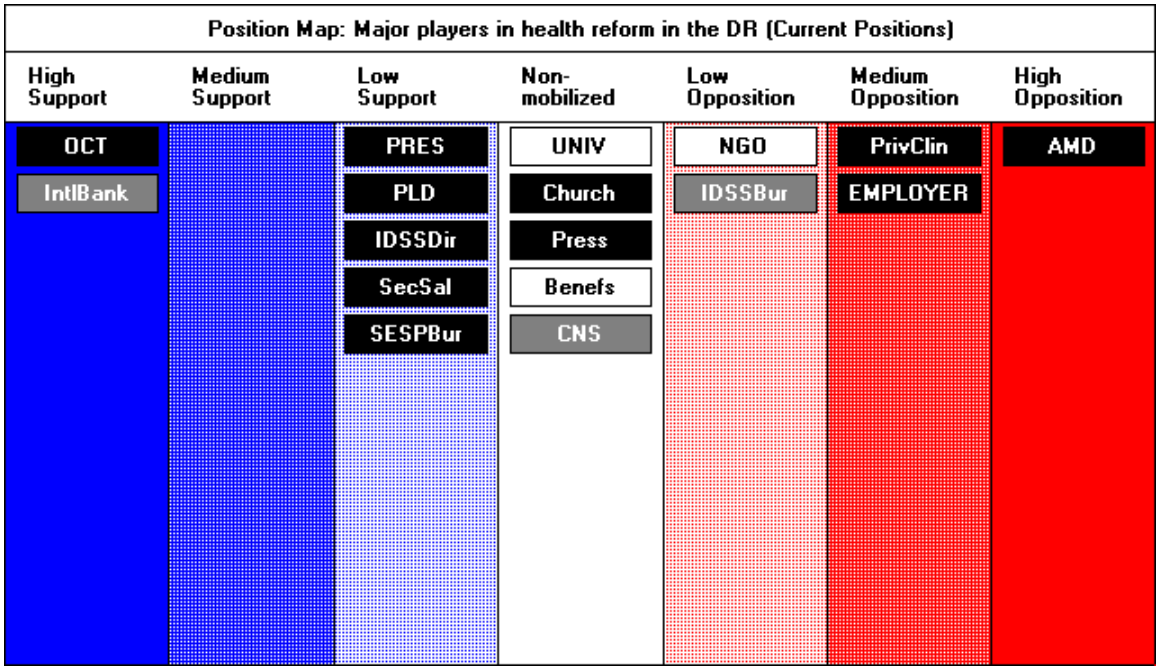
The critical role for public deliberation derives from our belief that there are no easy answers in health reform. On the one hand, people tend to build their understanding of health systems through established conceptual structures, which Thomas Kuhn (1962) termed "paradigms." These conceptual structures guide the definition of problems and the decision on solutions. Unfortunately, the established paradigms provide only limited analytic perspectives on health systems, with the result that both problems and solutions

tend to be artificially constrained, often masking real problems and preventing serious consideration of alternative solutions. In these situations, public deliberation can help identify uncertainties in facts and values, can contribute to a serious social process of grappling with what matters to a specific community, and can lead to innovative public policy design and democratic decisions.

Technical experts thus have ethical obligations in their consideration of political strategies. First, experts should be aware of the problems inherent to the reform process, particularly with regard to the potential for experts to conceal uncertainties of fact and value. Second, experts should be self-reflective about their own values and how their personal ethics shape recommendations for choices about public policy (as we have tried to be, in this book). Third, we believe that experts have an ethical obligation to encourage public transparency in helping societies confront public choices that involve uncertain facts and alternative values. We do not have evidence to prove that public transparency about these issues results in "better" social choices about health sector reform, but we do believe that this approach promotes a respect for individuals while building a tolerant yet supportive community.

These challenges make health sector reform more than just a technical process – it is a testing ground for social values and personal commitments. In constructing a plan for change, health reformers have to confront basic issues of principle, identity, and interest. Health sector reform can be as complex on a personal level as it is from the technical perspective. As they embark on the health reform voyage, advocates for reform need to understand how the political challenges and ethical dilemmas intersect with the technical issues, and how to address those intersections at both the personal and public levels.

Figure 5.1
Position Map for Major Players in Health Reform in the Dominican Republic in 1995



Key to power gradient: white box = low power; grey box = medium power; black box = high power

Source: Glassman et al., 1999.

Figure 5.2
Sources of Political Power

Tangible	Intangible
Money	Information
Organization	Access to Leaders
People	Access to Media
Votes	Symbols
Equipment	Legitimacy
Offices	Skills

References

Cobb, Roger W., Marc Howard Ross, eds. (1997). *Cultural Strategies of Agenda Denial: Avoidance, Attack, and Redefinition*. Lawrence, Kansas: University Press of Kansas.

Downs, Anthony (1972). "Up and Down with Ecology – The 'Issue-Attention Cycle'," *The Public Interest* 28(Summer):38-50.

Edelman, Murray (1977). *Political Language, Words That Succeed and Policies That Fail*. New York: Academic Press.

Glassman, Amanda, Michael R. Reich, Kayla Laserson, Fernando Rojas (1999). "Political Analysis of Health Reform in the Dominican Republic," *Health Policy and Planning* 14(2):115-126.

Johnson, Haynes, and David S. Broder (1996). *The System: The American Way of Politics at the Breaking Point*. Boston: Little, Brown.

Kingdon, John W. (1995). *Agendas, Alternatives, and Public Policies*. 2nd ed. New York: HarperCollins.

Kuhn, Thomas S. (1962). *Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Nandakumar, A.K., Michael R. Reich, Mukesh Chawla, Peter Berman, and Winnie Yip (2000). "Health Reform for Children: The Egyptian Experience with School Health Insurance," *Health Policy* 50:155-170.

Olson, Mancur (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge: Harvard University Press.

Reich, Michael R. (1994a). "The Political Economy of Health Transitions in the Third World," in Lincoln C. Chen et al., eds., *Health and Social Change in International Perspective*, Boston: Harvard School of Public Health, pp. 413-451.

Reich, Michael R. (1994b). "Bangladesh Pharmaceutical Policy and Politics," *Health Policy and Planning* 9(2):130-143.

Reich, Michael R. (1995). "The Politics of Agenda-Setting in International Health: Child Health Versus Adult Health in Developing Countries," *Journal of International Development* 7:489-502.

Reich, Michael R. (1996). "Applied Political Analysis for Health Policy Reform," *Current Issues in Public Health* 2:186-191.

Reich, Michael R. (1997). Review of *World Development Report: Investing in Health* (New York: Oxford University Press, 1993), *Economic Development and Cultural Change* 45:899-903.

Reich, Michael R., and David M. Cooper (1996). *PolicyMaker: Computer-Assisted Political Analysis*. Software and Manual. Brookline, MA: PoliMap.

Starr, Paul (1995). "What Happened to Health Care Reform?" *The American Prospect*, Winter, pp. 20-31.

Strom, Stephanie (1999). "Japan's Legislators Tighten the Ban on Under-Age Sex," *The New York Times*, May 19.

World Bank (1993). *World Development Report: Investing in Health*. New York: Oxford University Press.

World Health Organization (2000). *The World Health Report: Health Systems: Improving Performance*. Geneva: WHO, 2000.

Chapter 6
From Diagnosis to Health Sector Reform

1. Introduction

Once health sector reformers have decided which performance problems to focus on – that is, once they have decided on their strategic priorities – they have to go on to the next stages of the policy cycle. This means, first of all, figuring out the causes of the poor performance they are concerned with a process we call "diagnosis." Then they have to decide what to do about the situation – a process we call "policy development." In this chapter we want to discuss those activities and offer readers advice about how to do them so as to produce effective health sector reform.

With respect to diagnosis, the fundamental strategy is "work backwards." Keep asking "why" until you have discovered the causes of the poor performance you want to improve. The goal of that process, called a "diagnostic journey" – is to construct a "diagnostic tree" – an analytical device we describe in detail below – that links aspects of poor performance to those causal factors that might be changed or modified by policy interventions.

Once causes have been identified, the next stage in the reform cycle is policy development. This involves crafting a set of policies and programs that will alter the causes (or their impact) and thus improve health sector performance. Here we urge readers to pay attention to a simple general guideline: "Imitate but adapt." Because new ideas are hard to invent, imitating proven approaches has much to recommend it. On the other hand, because local conditions vary significantly, successful imitation involves adapting and adjusting ideas from elsewhere to local circumstances.

Our next guideline applies to how the previous two tasks are carried out. That is, "Process matters," and can have a major effect on the outcome of health sector reform. Whether or not critical actors and interest groups feel they have been consulted and heard will influence their "buy-in" to any reform plan. And that, in turn, can influence both the politics of getting reform measures adopted and the long run implementation of policy change.

In this chapter we begin by reviewing basic concepts and sources of data. Then we work through an extended example of a diagnostic tree, and discuss how to link diagnosis to the control knobs. However, thinking about performance problems one at a time, and in isolation from their interactions – including their lost implications – is likely to lead reformers to miss the larger picture. So having discussed the "one-problem" analysis, we then discuss how to shift one's focus from solving a specific problem to that of identifying critical causal factors that account for multiple problems and hence are an appropriate target for the more radical, multi-faceted efforts that constitute major health sector reform.

In the second half of the chapter we shift our focus to policy development. We first discuss some process concerns. Then we explore some screening criteria, by which to judge alternative proposals. And finally we discuss how to find and use evidence in both phases of the process – and conclude with a few final cautions and observations.

2. Using the "Control Knobs": Concepts and Evidence

As noted in Chapter 2, we have organized the body of theory and evidence about the behavior of the health care system under five "control knobs." Each control knob focuses on certain features of the system and the strategies that governments can use to alter those features in order to improve health sector performance. For example, in the finance arena governments can change what taxes are used to raise money for health care, or under payment how physicians are compensated, or under organization how it organizes hospitals etc. To design effective policy, a reformer needs to understand how changes in the control knobs will lead to changes in health sector performance.

At the diagnostic stage, the control knob framework can be a helpful organizing device. For at that stage the task is *not* just to identify causes but to identify causes, that can be *changed* by government action! The control knobs thus provide a menu or shopping list – a set of possible endpoints to the diagnostic journey that in turn can be the basis for potentially effective government actions.

Let us put this point another way. Identifying potentially effective changes in one or more control knobs is the end of the diagnostic task and the beginning of the policy development task. An analyst first works *backward* to causes. Next they have to look to policies to change those causes, and work *forwards* from causes to forecasts of improved health sector performance as a result of various policy changes.

In tracing out the likely effect of changes in the control knobs, we urge a process that is "evidence-based." There has been a large movement in medical practice, in recent years, away from clinical practice based on received wisdom to "evidence-based medicine," where decisions follow from careful study of the relevant scientific literature. Advocates of this approach argue that doctors often develop patterns of practice that are not scientifically justified. For example, some physicians routinely remove tonsils or perform hysterectomies when there is no real reason to do so. The same can be said of some health sector reformers. They routinely urge their favorite remedy – e.g. "privatization" or "decentralization" – without first determining the performance problems in a particular country and without a careful analysis of whether their favorite intervention will improve that performance.

To counter such uncritical enthusiasm we urge "evidence-based reform," where policies are based on careful analysis of problems, causes and likely effects. For example, suppose a newly independent country of Eastern Europe has identified declining male life expectancy (due to cardiac problems and rising suicide rates) as a priority performance problem. And suppose too, that someone urges that country to utilize the finance control

knob and develop a system of competitive private health insurance companies. Such a move would make no sense unless its advocates could tell a plausible story about how and why this "setting" on the finance control knob would do something to improve male life expectancy. (In fact we doubt that such a story exists.)

What kind of knowledge can we use to identify causes and predict consequences? What we would most like are well-designed studies that produced reliable statistical results linking policy to performance. However such studies are far less common in health systems work than clinical medicine. Health care systems are complex and varied, and there are few cases to study – i.e. relatively few countries. Hence it is difficult to get a large enough sample to deal with the many variables that are potentially relevant. In addition because data are difficult to collect, they are often incomplete, or unreliable. And it is almost impossible to run controlled trials on even one health care system – never mind a large number of systems – the way we can study a large numbers of patients.

The second most desirable sources of evidence are well-designed studies of demonstration projects that cover a number of both experimental, and non-experimental sites (to serve as control groups), and report comparative performance data. But these too are relatively rare. Comparative analyses of similar interventions in different countries are more frequent, and can be quite helpful in suggesting possible patterns of causality. Suppose, for example, we look at a selection of community financing initiatives, some of which went well and others badly. We then might be able to develop some ideas about the preconditions for the success of such interventions. More difficult to use, but still of value (and perhaps most common – especially for national level changes) are reports about the effect of a policy change in one country. It can be hard to disentangle causes of success or failure in such reports, but they can still be provocative or suggestive.

In addition to studies of various kinds, general theoretical arguments – about the effect of economic incentives say, or of political symbols – can also be brought to bear on a particular case. For example, if we pay doctors more to do X than Y, they often do more X and less Y. We do however have to be cautious because theoretical arguments may miss important local details – like whether hospitals will report accurately or whether actual local markets will be competitive. This leads to a third basis for diagnosis and prediction, namely an analyst's own experience and judgment as informed by detailed local knowledge. This can help someone see why a particular policy or program might not work in a specific country. For example, we might not opt for a regulatory scheme that worked well in the context of an efficient and respected legal system if our own national justice system is not especially reliable.

In subsequent chapters, on each of the control knobs, we will try to offer some "conditional guidance." That is, we will try to suggest both what the effects of various policies are likely to be, and how those effects can be influenced by local circumstances. But to do diagnosis and policy development most effectively, a reformer needs to be able to adjust even those conditional arguments to local conditions based on their own grasp of those conditions.

3. Developing a Health System Diagnostic Tree

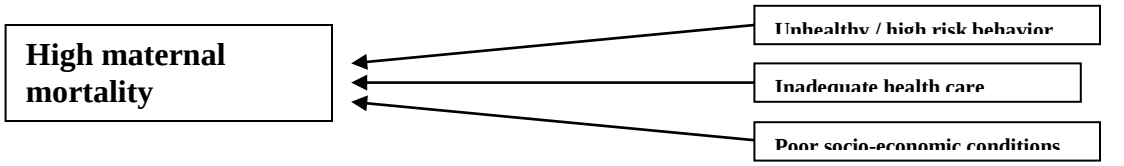
Before presenting a specific example of how to do a diagnosis, we need to calibrate our expectations about what such a process is likely to find. Given the complexity of the health sector, it will seldom be the case that A is caused by B, and only B; and that B is caused by C, and only C. Instead:

- There are likely to be several stages or series of causes
- Each effect may well be produced by more than one cause
- Each cause is likely to have multiple effects
- Causes and effects may inter-act and reinforce each other in various ways
- Not every cause will be manipulable by public policy
- Change may well require acting on more than one cause at a time – i.e. changing more than one control knob.

With this caution in mind, let's look at a specific example of a diagnostic journey. To do this, we will use an analytical tool we call a *health systems diagnostic tree*. Such a tree is similar to, but different from, two other kinds of tree diagrams the reader may be familiar with. One kind is a decision tree, which is used for analyzing decisions under uncertain circumstances. At each branching point in a decision tree, one of a set of mutually exclusive *events* occurs (either by chance or by the choice of the decision-maker). In a health system diagnostic tree, in contrast, at each branching point, an analyst represents various possible *causes* for a particular situation. And it may well be the case that several of the causes are in fact operating at the same time. There is no presumption that paths are mutually exclusive – unlike the decision tree. The health sector diagnostic tree is also different from the kind of tree diagram, implicit in the diagnostic process in clinical medicine. These trees also involve causality. But normally at each branch, only one out of a set of alternative causes is to be selected. The patient has *either* this disease *or* that one. In health sector diagnosis, more than one cause (i.e. more than one "disease") is generally operating at the same time.

Consider the following example. Suppose the government of a country decides that, compared to other similar countries, it has a noticeably high rate of maternal mortality – especially among the rural poor. Suppose further that on both political and philosophical grounds the government decides to make improving this rate a priority. We will work through this example, to see how the diagnostic process operates.

Figure 6.1.
Starting To Work Back* Along the Causal Chain



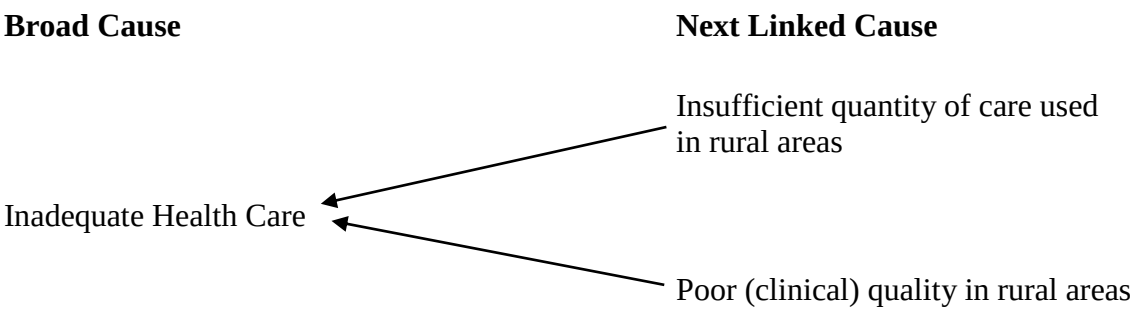
* In this and subsequent Figures, the arrows point in the direction of causation. But in the diagnostic process, we would build our analysis "backwards", from outcomes to causes.

The first step is to consider *broad categories* of causes. Figure 6.1 lists three possible causes that often contribute to such health problems in developing countries: inadequate health care, high risk behavior and poor socio-economic conditions. As part of health sector reform a country will not be able to address causes outside the health care system. But these should be still considered in the analysis, to keep a realistic view of the role that health sector reform can play in solving the problem.

One way to determine which links are significant, is to look at one's own situation compared to similar countries that do better. A possible causal factor that operates similarly in both nations is not likely to account for differences in their performance. For example, if two known behavioral risk factors (i.e.. hard physical labor, smoking or diet), are similar in our nation to those in countries that do better, than these causes are not likely to account for our high maternal mortality. A second line of analysis is to take advantage of any variations in outcome within a country, by region or socioeconomic group, for example. We can then look to see if there are any variations in potential causal variables that are parallel to these variations in outcomes. Such an association that does not prove there is a causal connection; but it should make diagnosticians suspicious enough to explore the matter further.

Suppose other countries with similar socio-economic conditions and behavior patterns do achieve better levels of maternal mortality. This should lead us to focus attention on health services as a potentially critical factor. The next question is described by Figure 6.4, namely, what aspects of the nation's health services are "inadequate." This second stage shows how a causal analysis reaches back through successive causes of poor performance (and causes of these causes). The health system diagnostic tree is just a way to keep track of, and map out, the causal chains involved.

Figure 6.2.
The Next Step Back Along the Chain

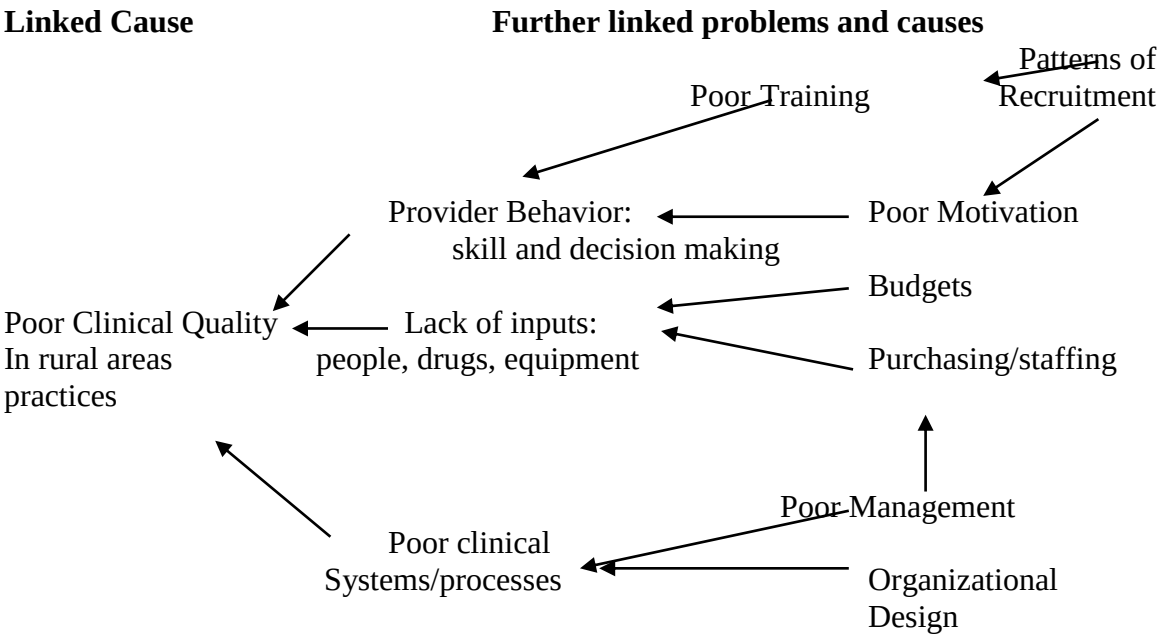


The first step in sorting out the relative contribution of the two causes described in the diagram is to look at utilization. What care do citizens actually receive? If, for example, antenatal care visits per capita or our rate of hospital deliveries are similar to those of better performing nations, quality, not quantity, is likely to be the problem. Lower utilization rates, in contrast, suggest a quantity issue. The death rate among patients who are admitted to a hospital (again versus international benchmarks) could also provide information on clinical quality. Since maternal death rates among the rural poor are of special concern, it would be important to look at these quality and quantity comparisons for that population versus the rest of the country. Moreover, as we noted initially, at each juncture in the tree, the options are *not* mutually exclusive. There could well be both quantity and quality problems, to varying degrees, in different regions of the country.

As we move back through the causal tree, we seek to identify factors we can alter through new policy initiatives. For example, suppose analysis reveals that poor clinical quality of care – especially in rural areas – contributes to maternal deaths. Then a deeper look at possible causes of that poor clinical quality, as illustrated in Figure 6.3., would be in order. The first two branches of that diagram are straightforward. First is there a problem with the providers? As we discussed in Chapter 4, evaluating skill and decision making will generally involve a look at clinical records. Investigating the care given in a sample of maternal death cases would be one way to proceed. Suppose this study does reveal clinical quality issues. This should provoke still further inquiry. First, were the right decisions made about how to care for patients? Alternatively our studies might reveal that while people tried to act correctly, they could not because the system lacks critical inputs – personal facilities, equipment, drugs and/or other supplies. Alternatively there could be problems at the systems level – poor communication, lost records, etc.

The third branch reflects this last possibility. As we discussed in analyzing clinical quality –it is often the way in which the care system is organized that accounts for poor results. Record keeping is poor or responsibilities unclear, etc. Indeed as we will see in Chapter 9 on Organization, much of the thrust of the total quality management movement has been aimed at making changes to such systems.

Figure 6.3.
More Steps Back Along the Chain

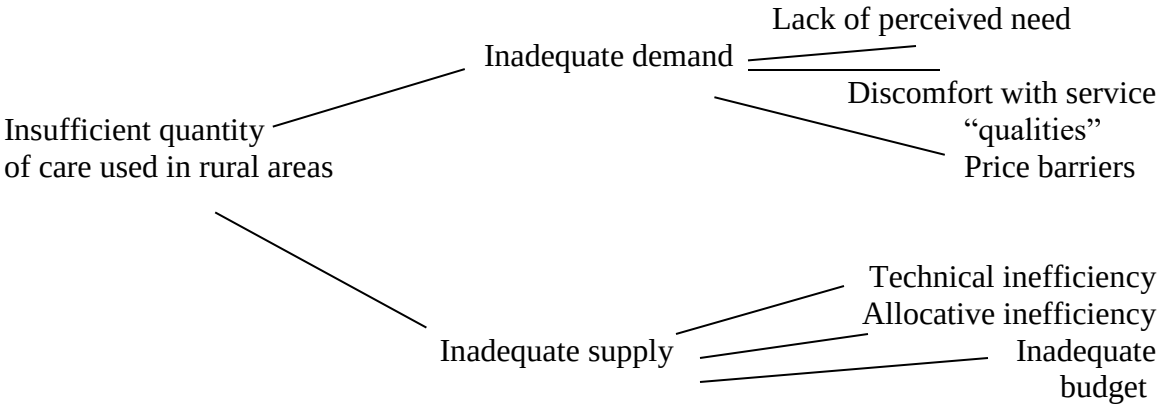


Now each of these causes gives rise to still further questions. Suppose wrong decisions are made. Is it because people don't know better (training) or are not trying hard enough (motivation)? If inputs are limited is it a matter of overall budget resources or how purchasing and staffing are carried out? If clinical systems and processes are faulty is it poor management within institutions, or the way care is organized across the system (or both)?

And going still further back, we begin to get to categories within our control knobs. If workers are not motivated or clinical systems not well constructed – because of poor management – why is that? Is it because there are no incentives on the organization to produce quality, or are there no incentives on managers to do better? Or is it that managers lack the skill and the authority to do better? Here we see the expression of the principles we noted up front – causes can have more than one effect and effects more than one cause. Obviously, each of these endpoints could, in turn, provoke further analysis. Why is training not better or why are doctors not more appropriately motivated? As we move backwards up the causal chain, we should be ever mindful of the imperative to base our work on evidence, not simply speculation.

Alternatively, suppose the assessment depicted in Figure 6.2. reveals that an insufficient quantity of services being used in rural areas is a critical cause of our poor performance. The further study then required is shown in Figure 6.4.

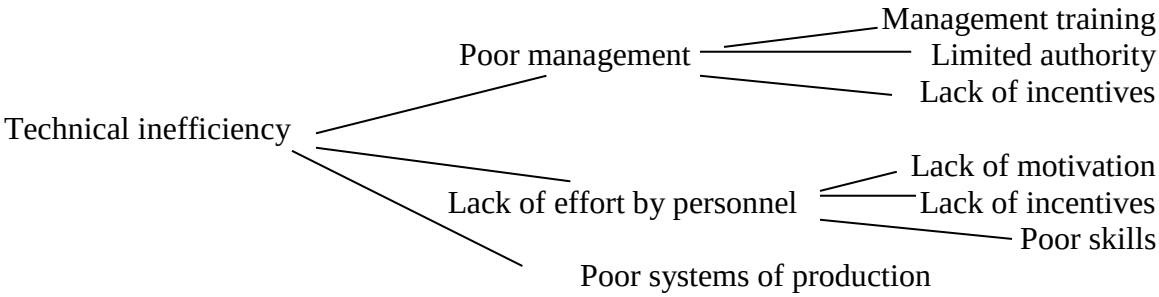
Figure 6.4.



If the quantity of care being utilized is insufficient, the explanation could lie on either the supply or the demand side. The simplest cause to explore involves simple physical availability. Are there clinics and hospitals with staff, supplies and the capacity to deliver more services? If that is the case than either patients are not using the services that are there, or some barriers are preventing physical availability from becoming effective availability (as we defined those concepts when discussing "access" in Chapter 4). If services are not even physically available, asking "why" leads to various possibilities. Perhaps the budget for these activities is too low. Alternatively, if spending levels compare favorably with other regions or other countries, the lack of services could be due to either technical inefficiency (i.e. high unit cost leading to lower volume) or allocative inefficiency (the wrong mix of services is being produced from the perspective of cost-effective performance). And if any (or all) of these turn out to be the case we are coming closer to the point of identifying causes that might be changed by acting on the various control knobs.

On the other hand, if services are physically available but unused, the issue is on the demand side, First there might be a lack of perceived need – that is, the local culture may strongly favor non-institutionalized delivery. Alternatively service "qualities" may make the service unattractive. So too might the presence of user fees (either explicitly to the institution or in the form of informal payments to providers or the requirement on patients to pay for drugs and supplies). Clearly then supply and demand are not fully separable because low demand may be due to the nature of the supply.

Figure 6.5.



To see this, let us trace out a yet one further set causal connections as in Figure 6.5. In particular if the system is technically inefficient, (high cost) why is that? Is it poor management, lack of effort from workers, or failures in the system of production? Suppose, for example, studies reveal that while clinics have adequate staff, in theory, staff only work a few hours a day – and spend the afternoons in private practice. Hence reported caseloads per doctor are low (and average costs high). To correct this the diagnostic journey has to go even deeper and look at staff incentives, organizational culture, systems of accountably (or their absence), how managers are selected and rewarded etc. etc. For this analysis will reveal the ultimate end of the journey namely to identify potentially changeable features of the situation. However, be warned. As noted above, multiple, interacting cases may well be a work and multiple interacting policies may be required to produce better results.

Notice that as we have done this analysis, quite a number of the intermediary criteria identified in Chapter IV have made an appearance. Varies aspects of quality, efficiency and access have played a role as part of all potentially relevant causal chains. Notice, too, however that categories like “technical efficiency” or “clinical quality” do not appear at the ends of the branches of the diagnostic tree. Rather as intermediary criteria they appear in the middle of the journey and their appearance itself requires further explanation.

As noted, the pathways in these diagrams are not mutually exclusive. Going all the way back to Figure 11.4, we could have both a quantity and a quality problem. The same causal forces could be producing both. Hospitals might be paid in a way that does not generate any incentives for performance. That, in turn, could produce both poor quality clinical care and also technical inefficiency that limits the quantity of care. Moreover if we did the full tree in all its branches we would see that many of the factors which contribute to poor quality also are potential causes of technical inefficiency. That is because poorly performing organizations are typically inefficient in the production of everything: quantitatively as well as qualitatively. Moreover, any one deficiency could be the result of several interacting causes. Perhaps our hospitals both lack incentives to do well and are also operated by managers with little authority to make improvements. These factors together could each make the impact of the other greater still.

As we develop the causal analysis further back along the causal chain, we postulate more and more specific causes that in turn begin to allow us to identify specific remedies. This process of diagnosis leads us back to the health system control knobs.

4. Linking Diagnosis to the Control Knobs

We have argued all along that the "control knobs" provide us with a framework for considering possible interventions to improve health sector performance, once we have gone on a diagnostic journey. In effect they provide a list of options and as a way of illustrating their use we can ask how could we use them to help with the high maternal mortality example we have been exploring?

- Financing: If we do have a quantity of service problem, can we tap new or additional sources of funding to expand the budget for this service? Alternatively can we alter how we spend existing funds to increase allocative efficiency
- Payment: If we discover we have a problem of motivation – at the level of the organization or the individual provider, can we change how we pay, or the rates we pay, to improve both the quantity and quality of services?
- Organization: If we are not organized to facilitate economics-of-scale and coordination, if patients experience cultural barriers, or have preferences for certain types of providers, what changes could we make in who offers various maternal health services to overcome these problems? If the issue is technical efficiency and/or clinical or service quality, can we change how providers are organized at the micro-level in ways which would increase their performance?
- Regulation: Are there regulatory devices we can deploy to increase the quality and quantity of services or to increase their utilization?
- Individual Behavior: If there are important causes on the demand side, could we seek to influence mothers to make more use of the services that are available?

Notice, to get significant change in the performance of our system with respect to maternal mortality, it is likely that we will have to take several actions simultaneously. For example, suppose we reorganize at the micro level to give managers increased authority – in order to improve efficiency. This is likely to have much less impact, if done in isolation, than if it is done in conjunction with payment reform (e.g. incentive contracting) of a sort that gives managers a strong reason to improve performance. Similarly, raising more money (via financing) to push into a system of care that performs poorly, may have very little impact without complementary changes in payment and organization.

Similar kinds of inter-action effects are likely to operate when it comes to producing changes in individual behavior on the demand side of this problem. Suppose we are trying to convince skeptical patients to increase their utilization of our reorganized maternal health services. That reorganization has to deliver changes in clinical quality and service "qualities" that will produce positive experiences for those who utilize the service. Indeed to truly make a difference we may need to raise more money (finance)

and deliver it via new incentive schemes (payment) to restructured providers (organization) – even as we seek to convince patients that the new services are worthy of their patronage (individual behavior).

Clearly the relative importance and relevance – as well as the specific content – of such interventions will depend on the details of a country's situation. For example, if services are largely provided in facilities directly operated by the Ministry of Health, it will be easier to influence their internal organization than if they are offered by independent practitioners. In the latter case it is likely that incentive payments and/or regulatory constraints will have to play a larger role. Similarly a nation with a public delivery system manned by reasonably well motivated employees, who are hamstrung by restrictive budgetary practices and poor management, will need different organizational changes than one in which staff are cynical and uninvolved.

5. From Specific Problems to Major Health Sector Reform

So far we have been discussing the process of health systems diagnosis and policy development as if it either does or should focus on one specific performance problem at a time. Such an analytical perspective is likely to provoke piecemeal, programmatic initiatives – no new maternal health program for example or an expanded effort at childhood immunization. But as we have noted before, that is not always the situation that develops in real reform efforts. Instead there are liable to be a number of performance problems identified in any given country at any one time. Moreover the diagnostic trees for these various specific problems may well reveal that certain features of the health care system appear again and again as important causal factors. Moreover dealing with multiple problems may well require multiple, significant interventions for their solution. And it is thus this kind of analysis that can lead a country to major health sector reform efforts.

In this connection we have to return again to the key role cost is likely to play in all problem definition efforts – as described in Chapter 4. A nation unhappy at its health status – and yet facing cost constraints – will often be forced into a complex and multi-part analysis as it seeks to unpack the causes of its difficulties. Everything from technical inefficiency (due to poor incentives from the payment system and poor management due to faulty organization) to allocative inefficiency (due to poor resource allocation from the financing system) to poor quality and limited access is likely to emerge. And program-by-program fixes will not be possible – exactly because there is not enough money to pay for a whole series of uncoordinated new initiatives. Again, a multi-part reform, where major changes are made to several control knobs, is likely to be the answer.

Consider the example we discussed earlier in the chapter – an effort to diagnose the causes of a high rate of maternal deaths, especially among the rural poor. As a part of that process (see Figures 6.4, 6.5 and 6.6) we explored the possible causes of poor quality of care and low use of services in rural areas. Now suppose a country decided it was also interested in the excess of childhood deaths in the same population and/or low overall life

expectancy among these groups. Then much of the diagnostic tree would look similar if not identical! Poor service quality and lack of utilization would appear again and again. Since the same budget-based public system provided care in all these arenas, an appropriate response to such a situation would likely be a multi-pronged attack, using several control knobs, to improve service and increase utilization in the rural areas. This could lead to a complex program of health systems reform, designed to improve performance without raising cost. For example, substantial authority could be decentralized to local communities, who would also be allowed to set up local insurance schemes (community financing) to generate some added funds. And those communities could also be allowed to contract with NGO's or private practitioners on either a capitated or incentive basis to operate local health centers – to get better results from the available resources. Such a scheme, involving a significant number of significant changes, would involve risks and demands that could not necessarily be justified by trying to fix one performance problem in isolation.

Here is another example. Suppose a nation is concerned about a variety of differences in health status and risk protection between those working in the formal sector (government and large employers) on the one hand, and those who are self-employed or who work in the informal sector on the other – as revealed in household surveys. Suppose also that latter group has mobilized politically to express their dissatisfaction with the current system. Moreover suppose the diagnostic journey for each of these problems bring us again and again to the fact that only those in the formal sector have access to functioning social insurance schemes – while other segments of the workforce largely pay for care out-of-pocket. In that case a nation might well decide on a major change in the financing system – e.g. on trying to establish a new unified social insurance system – if the overall level of national economic development was high enough to support such an initiative. This would both increase access to care among workers in the informal sector and give them the risk protection they now lack. And it might join this with efforts to institute selective contracting by the new fund –together with reform of the hospitals themselves to give them more autonomy and their managers more authority. Again such a multi-faceted reform enterprise would be based on a series of specific diagnostic journeys. But the overall portfolio of changes is a response to a complex set of intersecting phenomena.

It is possible to be generous and/or optimistic and suggest that this kind of multiple diagnostic analysis is implicit in the arguments of those reformers who advocate substantial policy changes without linking their proposals to any particular expected performance gains. However in our experience that is not the typical case. The failure to offer any systematic diagnostic analysis is more likely to reflect badly designed or poorly thought through reform ideas. Indeed our argument is that a complex reform program, designed to improve performance on multiple dimensions, requires more, not less, systematic and systemic analysis than simpler and more modest reform alternatives. The multiple diagnostic trees need to be drawn and the potentially complementary and conflicting consequences of reform programs need to be explicitly analyzed.

6. The Process of Policy Development

Before we say more about the substance of health sector reform, we need to say more about the process. Experience shows that process may matter as much as substance, and indeed that these two are deeply intertwined. In particular, involving interest groups in the process of policy development serves at least four functions.

First, participation by those affected allows policy makers to hear, and take into account, interest groups' real concerns. This can improve both acceptance and effectiveness. As we discussed in Chapter V, one way to assemble a support coalition is to compromise with key constituencies on features of a reform that matter more to them than they do to reformers. And allowing constituencies to participate in the policy development process is one way to learn about those concerns and take them into account.

Moreover being aware of such concerns can sometimes improve a policy from the point of view of successful implementation. For example if a hospital reporting system is too costly, it may be resisted even after it is legislatively imposed. Or perhaps it is simple not practical given hospitals existing data systems and administrative capacity. Learning about such issues in the policy development phase can be very helpful.

Second, participation can increase the acceptance of a plan because it increases the legitimacy of the process that developed it, both psychologically and philosophically. Being heard can "co-opt" participants – who feel that at least their concerns have been attended to. Exclusion, in contrast, operates both emotionally and prudentially to generate hostility and suspicion. And in a democratic society allowing interests to be heard is likely to increase the legitimacy of the process (and its product) not only among those who do participate, but among uninvolved third parties as well. For participation meets certain cultural and/or ideological requirements.

Third, participation can serve to educate interest groups about the concerns of, and pressures exerted on reformers by, other constituencies. For example the doctors learn what the hospitals, or unions, or consumers are demanding. This gives reformers both an excuse and an explanation for why they cannot necessarily respond favorably to any one interest group's demand.

Fourth, participation educates the participants about the detailed content of the proposals that ultimately emerge. If for example local governments have been involved in crafting a health sector decentralization proposal, they are more likely to know what exactly it will mean for their own responsibilities. This allows the relevant groups to understand the logic of reform, makes them better able to predict the consequences to them of it, and hence lowers their uncertainty. And at least some of the resistance to change and reform is generated by anxiety and uncertainty over what it will entail.

However it is also true that participation can go too far. There is a great deal of evidence that the skill and ability of various interest groups, varies widely when it comes to influencing policy development. Groups with more resources and expertise tend to

have more impact. Thus the balance of forces in a participatory process may be quite different from the relative importance reform advocates place on the concerns of the various interest groups.

As we have suggested several times, when it comes to health sector reform, this imbalance often lends a conservative bias to the process. Existing organized groups – especially providers – can overwhelm the inexpert and unorganized. Poor, rural consumers for example, seldom have much of a voice. Moreover the new institutions or organizational forms that might be created by reform do not yet exist to advocate for themselves.

Unfortunately, also, participatory processes can too easily degenerate into the demand for unanimity. This gives entrenched interests a veto – which can cripple any real hope for change. It can also lead to reports, and/or legislative language that is intentionally vague – exactly for the purpose of allowing those with diverse purposes to sign on to the compromise. This outcome only defers hard choices to the implementation stage – where again well-organized interests can often prevail (as we discuss in the chapter on Regulation below). Indeed powerful interests sometimes strategically support vague language exactly to defer important issues to a less-publicly-visible implementation phase, where they expect to have more influence.

All of these observations suggest that participation by interest groups raises, not lowers, the need for leadership by would-be reformers. Reformers can do much to shape any participatory process by defining its procedures and its membership. They can formulate questions, set agendas, specify alternatives and analyze implications. They can, and should, honestly explain the limits of the conversation and specify what options and spheres of action are, and are not, on the table.

Moreover those in charge of managing a diagnostic or policy development process need to understand that large groups of people are not capable of serious analytical work. At the front end they can "brainstorm" and generate new ideas. At the back end they can make selective criticisms of proposals on the table. But the hard work in the middle, of going from "brainstorm" to a detailed and developed policy option, together with an analysis of its implications, is not the work of a large meeting. Instead it requires organized and disciplined staff resources. And again, sophisticated managers know that the background, training, politics and biases of those doing the staff work can affect the result. This is especially so when data is scarce or inconclusive – as it often is in health sector reform (a point we return to shortly).

Finally we need to say a word about the process of generating policy alternatives. Much research on human problem solving and cognition supports the notion that new ideas are extremely difficult to generate. People in general have a difficult time thinking "outside the box" of their familiar concepts and paradigms. All of this implies the lesson we announced at the beginning of the chapter: "Imitate but adapt."

Imitation can take various forms. Ideas from health sector initiatives in other countries are one obvious source. Another possibility is ideas developed in other policy areas that can be adapted to the health context. For example, if decentralization has worked well in education, perhaps these ideas can be applied to health services. If contracting out in the highway department has improved the quality of work, there, again, maybe there are lessons to be learned.

General theoretical arguments can provide a third source of new ideas. Perhaps ideas about competition in economics or tort liability in law can suggest reform initiatives. Now here we do need to issue some words of warning. Economists in particular are easily overly enthusiastic about ideas rooted in simple formal models. Hence especially for such ideas, but also for any other new initiative – it is important to observe the "but adapt" part of our guideline "imitate but adapt."

To summarize this part of our argument then – process has an impact on content, politics, and implementation. Participation is not just a ploy. It is both a learning and a teaching device. It can improve a policy substantially and have a major impact on its acceptability. But it can also get out of hand and make the life of a reformer extremely difficult. Thus the process of policy development needs to be strategically managed. As with everything else about health sector reform, foresight, careful analysis and self-critical thinking can make a substantial contribution to the prospects of success or failure.

7. Screening Tests for Policy Interventions

Once plausible policies have been identified, reformers need to do a serious feasibility and implementation analysis. That analysis has to take account of the likely politics before and after a program is adopted; of local customs, capacities and institutions; and of the available financial and human resources. Such analysis in turn has to lead to a plausible story about how the proposed set of policy interventions will act "forward" down the diagnostic tree, to produce the desired performance changes. Unfortunately this is not always easy to do because our knowledge of the health sector, and of the social and economic systems connected to it, is far from exact and complete. Hence we are often not able to fully predict the impact of all possible reforms on all possible outcomes. However there are certain characteristics of reform proposals themselves that have general predictive value. We have alluded to these several times but now we want to describe these tests more specifically.

7.1. Implementability

We have argued that the critical test of a policy is what actually happens when it is implemented. Thus reformers have to explore whether or not their country has the institutional and social prerequisites needed to support a proposed reform. For example, do the government agencies that will have to carry out the program have the necessary technical skills and managerial sophistication? Can the existing systems of data collection reliably produce what will be needed? How reliable are the courts and police

when it comes to enforcing proposed regulations? There is also the matter of influence from special interests. What role will various groups – from tribal leaders, and political parties, to the medical society – have in the implementation process, and will this influence undermine a program's real impact? All of these factors will help to determine what any policy will actually accomplish.

Social and cultural conditions and norms will also have an impact on implementation. How deferential to authority is the population? How willing is the government to use coercion to get its way – and how legitimate will such efforts be in the eyes of the population? Is rule compliance itself widely valued – or widely flouted and disrespected? Will people lie, or tell the truth when asked to report on their own behavior?

As we stressed in discussing the policy cycle in Chapter 2, wise policy designers carefully consider such issues in the process of policy development. They don't propose payment systems that will tempt the unscrupulous or regulations that require unavailable data for their monitoring. Indeed, as we review the individual control knobs, issues of implementability will appear again and again. Wise reformers do *not* presume that an idea from Denmark can be replicated in Uganda or that a scheme that was effective in the U.S. will work in Bosnia. Instead they will look for schemes that fit the cultural milieu and organizational capacity of their environment.

7.2. Political Feasibility

Just as those doing policy development need to look forward in the policy cycle to implementation, they also need to look forward to the process of political decision making, which they have to go through to get their reform accepted. Political feasibility itself is both contextual and agent-relative. Whether something is feasible will depend on the resources, commitment, and skill of those seeking to get a plan adopted, as well as on the other forces at work in the political process.

In Chapter 5 we discussed the ideas of political mapping and political strategy, and showed how these effect political feasibility. This analysis makes it clear that feasibility is not a yes or no judgement. It depends not only on who starts out in favor or opposed to plan, but also on who is prepared to expend what resources, and take what risks, to get a plan adopted. It depends as well on the political skill of a plan's proponents. Even then, the outcome is almost always uncertain. Hence feasibly is best described as a probability, conditional on the level of effort to which proponents are willing to commit.

Thus reformers need to ask themselves – how hard are they and their allies prepared to fight to get a particular scheme adopted? Their real choice may be between a plan that it is easier to get approved – and one that is politically more difficult to adopt, but which promises more impact on core performance criteria. Thus a political cost-benefit calculation, based on the particular situation (political resources, goals, philosophy, etc.) of reform advocates, is implicit in the notion of political feasibility. For the critical tradeoff may well involve personal costs to the reformer versus benefits to the society as a whole.

7.3. Political Controllability

A final consideration in policy design is whether or not new arrangements and institutions will be subject to effective democratic political control. Such control depends on many features of the situation. First will the results of the policy be publicly knowable, i.e.; are the consequences measurable and those results transparent? For unless data are both reliable and available, outcomes cannot be measured. And what cannot be measured cannot easily be managed. Second, once we know the results of a policy, are there mechanisms in place to change it in response to information about its performance? Finally, are those who have the power to make such changes themselves politically accountable? That is, if citizens are unhappy, do they have channels of political action available to effectively express their dissatisfaction and through which to push for further reform?

The general argument for such controllability is based in part on the likely imperfection of any reform. As we argued in Chapter 2, the reform cycle is a cycle. New and /or unanticipated problems appear. If we create new institutions and arrangements that are difficult to change, our initial efforts can easily become an obstacle to progress in later rounds of reform. Moreover as egalitarian liberals we believe that respect for individual autonomy implies that people ought to be able to have an impact on those public policies that have a large impact on them. Political controllability, we believe, will also generally produce better outcomes in the long run. Consumers, patients, and citizens who have a political voice are more likely to demand, and get, better performance from the health system.

However this test is also controversial. There are some within the health sector who want to limit political control. For example, in transitional economies in Eastern Europe, physicians favor the establishment of social insurance funds isolated from the general budget process. This is because they do not trust political institutions, pressed by many competing demands, to produce the funding levels for health care which they desire.

We are not naïve about the consequences of political controllability and accountability. These can be undesirable in specific cases. For example, transparency can increase the influence of special interest groups--who use information about what is occurring to exert pressure upon the legislative process. Such arrangements also can make it more difficult for legislators to support proposals they believe are in the national interest – when those same proposals undermine the parochial interests of their constituents.

On the other hand, benevolent dictators or powerful technocrats will only produce "good" results if one happens to agree with their substantive values – a risky long run strategy. And our own inability to fully predict consequences is a strong argument against trying to prevent revisions of our work. Thus we come down if favor of trying to build some public, self-correcting properties into the institutions we create as part of health sector reform.

8. Finding and Using Evidence for Diagnosis and Therapy

Evidence-based health policy requires…evidence! And that is something that many lower and middle income countries do not have in abundance. Ironically though, even when data and information are available, they are often not well utilized for policy analysis. So we need to be concerned both with the availability of different types of critical information as well as its effective use. The recent report of the Global Forum for Health Research noted that "…like other global public goods, global health and global health research suffer from insufficient investment…" (2000).

It is clearly beyond the scope of this chapter to provide extensive guidance on the many facets of health and health systems research that could support national efforts of diagnosis for health reform. But to help get things started, here are some brief suggestions about the more important types and sources of data.

Some type of measurement of performance of at least one of the three main criteria – health status, financial protection, and customer satisfaction – is essential for effective diagnosis. In the 1990s there have been important advances in collecting, assembling, and disseminating data on health system performance. For the lower and middle income countries, this has included the reports of the International Commission on Health Research for Development, the World Bank's 1993 report *Investing in Health*, and most recently WHO's Health Systems: *Improving Performance* (2000). The annual publication of the Organization for Economic Cooperation and Development's *Health Data* provides an invaluable source of data on OECD member countries. The national reports on central and eastern European countries, produced by the WHO and the European Observatory on Health Care Systems (the "Health Care Systems in Transition" or HIT series) are also very helpful. Each of these major reports includes major statistical annexes tabulating a variety of indicators of outcomes and health system characteristics.

That said, diagnosticians are cautioned that the availability and quality of the evidence in these global references is very uneven and often suspect. Local data sources are often available which may be both more trustworthy and more trusted by decision-makers. The national response to the World Bank's 1993 report and WHO's 2000 report was quite mixed. Individual country figures were sometimes extrapolated from regional averages, which can be misleading for countries with better or worse performance than the average in their regions. National health expenditure information in global databases is sometimes based on significant "data-filling" where local sources may be available but not known to the international compilers of this information.

Of the three main performance criteria, evidence on health status is by far the most developed. Traditional health status measures, like infant and under-five mortality and life expectancy have now been augmented with new composite measures of disease burden like DALYs and DALEs (WHO, 2000). Many countries now collect significant national health status information using standardized survey instruments such as the

Demographic and Health Surveys (DHS) and the Living Standards Measurement Surveys (LSMS). These can be supplemented with data from well-done smaller local studies, which may provide additional details on specific health problems and population groups.

Evidence on the other two criteria – financial protection and customer satisfaction --is more limited, especially in lower and middle income countries. For example, evidence on the distribution of the financial burden of health care may be available from national health accounts studies or household surveys. But the concept of protection implies concern with the *result* of the financial burden in terms of household financial status. And here little general data are available. As noted in Chapter 4, this depends not only on the health system but on other socio-economic factors as well. Measurement of customer satisfaction with health systems has been extensively done in the OECD countries using polling methods, but is not always easy to interpret. Little evidence on satisfaction is available from the lower and middle income countries. The recent effort by WHO in the report on *Health Systems: Improving Performance* is worth noting, but the quality of this information is often not sufficient for it to be used for diagnostic purposes.

Individual countries may have some evidence on both financial protection and customer satisfaction from national surveys and smaller-scale studies. We strongly encourage diagnosticians to assemble and review this information and improve upon it if they can.

When it comes to many intermediate outcomes and factors representing determinants of health system performance the good news is that there is much more evidence available. Some of the key data sources include:

- National health accounts, health care costing studies, and financial reports of insurers and other payers on financial resources, efficiency, and costs;

- Household surveys and health information system reports on access and utilization;

- Quality studies and health information system reports on clinical quality of care;

- Government statistics, professional organization statistics, and facility-based studies on availability of facilities, manpower, equipment, supplies;

- International pharmaceutical market data, government and private market statistics on pharmaceutical availability and use.

We know from experience that there is a lot of relevant information available for health systems analysis in almost every country. But we also know that this information may not be easily accessible at first and that it often needs to be used with caution.

Here are some of the ways you can make more efficient use of the available evidence are:

- *Know the literature:* It is increasingly easy to access up to date information through international publications and the Internet. Take the time to identify the relevant literature and become familiar with it. It is unlikely that you are the first person to confront or think about the problem that you are trying to solve;

- *Get advice*: National and international expertise is often available to help you sort through the literature and experiences of other countries. The major international organizations can be helpful as can international networks like the European Observatory or the Alliance for Health Systems Research. But…*caveat emptor* (let the buyer beware). Be skeptical of advice as well, especially advice that comes from those with something to sell, like a costly computerized information system;

- *Benchmarking*: Decision-makers often want to know the answer to normative questions like "how much should we spend on health care" or "should I allow private health insurance to develop". These questions don't have simple answers. But one method for using international evidence is to look for associations between key indicators of intermediate outcomes and causes and the final outcomes of interest;

- *Rapid assessment*: Many questions do not require a major research project in order to develop a plausible account. Just because detailed evidence on an important point isn't available doesn't mean there aren't quick and relatively inexpensive methods to get plausible estimates;

- *Support good research*: Despite the last point, there may be real problems with "quick and dirty" studies and assessments. Sound applied research can provide more valid and reliable evidence. The bigger the stakes, the more costly it will be to make mistakes. In our experience, most lower and middle income countries grossly underinvest in health systems research. But since such research takes time, planning and resources a good deal of foresight and toward thinking is required if you are to have the data you need when you need it.

9. Diagnosis and Policy Development: Some Final Observations

To do a good job of natural health system diagnosis, the health system reformer must be well prepared, curious, careful, and willing to learn from experience. The "science" of health sector reform is still in an early stage of development. Here are four final points we urge you to keep in mind as you do your own work.

- *Work backwards up the causal chain until you can identify potentially manipulable variables*: The job of health sector diagnosis to go from symptoms to causes, from outcomes and outputs to that which is producing them. This process is not always obvious. Policies can produce unintended consequences. Programs

can be badly managed. Individuals can claim to be doing one thing and in practice, be doing something else. An open and skeptical mind, and a high degree of energy and curiosity are often required;

- *Do not jump to quick conclusions*: It is tempting to decide early on what the problem *really* is based on ideology or preconceptions. Unfortunately, it is not uncommon for a committed true believer to short-circuit the diagnostic process. This can be unfortunate since identifying the wrong cause can lead to badly designed policies. This will not only divert valuable energy and waste a valuable opportunity, but it can even discredit reform generally;

- *Be scientific, not judgmental*: Diagnosis can go astray when evaluative judgments cloud scientific analysis. If institutions or individuals behave in ways we disapprove of, we are quite likely to decide that they are the problem. But, in fact, not every bad actor is important. Moreover, bad behavior may just be the result of larger forces. Thus, it is important to remain dispassionate in the diagnostic process. There will be plenty of time (and need) for passion, once you are trying to mobilize political support for a particular policy. Meanwhile, clear, self-skeptical analysis is called for;

- *Use numbers*: We do not believe that everything important to health sector diagnosticians can be reduced to numbers. Still less do we believe that every country will have – or can afford – all of the good quality data an analyst might like. And ignoring problems that cannot be documented in numbers can be very foolish. Disaster planners know that after earthquakes, the *silent zones* that they don't hear from can be the most badly affected. The same is true for health systems performance. We may get no data from the poorest and worst off areas. Yet, for all of that, data can be very helpful. Data can be a check on bias and premature judgement. Data can help support a scientific attitude. Whether it is DALYs, or other health status indicators, whether it is National Health Accounts or surveys of citizens' attitudes, data can be a great leveler of prejudices. To use data wisely you have to know where it comes from, its limitations, omissions, and assumptions. But that only increases the need for making the requisite investment in such understanding.

References

Berman, P. and T. Bossert, "A Decade of Health Sector Reform in Developing Countries: What Have We Learned?", Data for Decision Making Project, International Health Systems Group, Harvard School of Public Health, Boston, MA., 2000

Bradley, G. W. *Disease, Diagnosis, and Decisions*, John Wiley and Sons, New York, 1993

Global Forum for Health Research , "The 10/90 Report on Health Research 2000", Geneva, 2000.

Lusted, L.B., *Introduction to Medical Decision-Making*, Charles C. Thomas, Springfield, Illinois, 1968.

Organization for Economic Cooperation and Development (OECD). OECD Health Data (CD-ROM), Paris, 1999

Sackett, D.L., W.S. Richardson, W. Rosenberg, R. Brian Haynes, *Evidence-based Medicine: How to Practice and Team EBM*, Churchill Livingstone, New York, 1997.

World Bank, *Investing in Health*, Washington, D.C. , 1993.

World Health Organization, *Health Systems: Improving Performance*, Geneva, WHO, 2000.

Introduction to Part II
Overview of the Control Knobs

1. Introduction

In the previous chapter we introduced a new term – the health system "control knob." In the five chapters that follow we devote considerable attention to each of the five control knobs: finance, payment, organization, regulation, and behavior.

This brief introduction to Part II of the book has three purposes. The first is to give the reader a clearer idea about what we mean by the term "control knob." The second is to provide a brief introduction to how the control knobs can be used to address the performance problems discussed in Chapter 4. We will return to this theme again in each control knob chapter and at the end of the book. The third is to call the reader's attention to various features of the health sector that are *not* included under any of these five headings.

2. What is a health system control knob?

The term "control knob" is a metaphor. Think about some analogies of complex, purposeful activity intended to produce some result or benefit. Consider for example, the pilots of a large aircraft adjusting different controls to the right settings to achieve, altitude, speed, fuel economy, and safety to reach a desired destination. Or, consider the staff of the control room of a steel mill, setting the rates of different steps in the production process, the temperature of the blast furnaces, the quantities of raw materials processed, etc. In each of these examples the managers of a complex system must select the settings on a range of factors that determine how the system operates and what it produces.

A health system control knob, then, has two necessary characteristics. First, it must be something that *can be adjusted* by government action as a result of policy choice or policy decision. Second, adjustments or changes in the control knob must be a *significant causal determinant of health system performance* via causes affecting the intermediate criteria and ultimately the core performance criteria of the health system. In other words, our control knobs describe discrete areas of health system structure and function that matter fundamentally for health system performance and are subject to change as part of health reform. Not everything that matters for health system performance is a control knob, if it cannot be changed as part of health reform. And not everything that health reform can change is a control knob, if it does not fundamentally determine health system performance.

This idea of a health system control knob as having settings that can be changed to bring about changes in health system performance distinguishes our approach from simply describing health systems. For example, in *Health Systems: Improving Performance* (WHO, 2000), WHO describes health systems in terms of "…what they

do—how they carry out certain *functions*—in order to achieve anything." [p. 24, emphasis in original text]. This approach is useful for describing different ways in which health systems are organized and for comparative description of health systems. But our approach focuses on developing theory and evidence-based approaches to changing health systems to improve their performance. Our choice of critical variables for health system reform (the control knobs) represents the factors that cause a health system's results and can be used deliberately to change those results.

3. Using the Control Knobs to Improve Health Systems Performance

Each of the next five chapters explores in more detail how each control knob affects health system performance and also provides some conditional guidance on how each control knob can be used to achieve different objectives. In the concluding chapter, we revisit our framework linking diagnosis to health sector reform to consider some of the important issues in developing programs of fundamental health system change. But here, as we start on a more detailed discussion of each control knob, we want to highlight a three important cross-cutting points, as follows:

- Each control knob can be adjusted individually, but health reform almost always involves using more than one control knob.
- Control knobs affect health system performance directly, but they also affect other control knobs, which affect system performance. There is interaction across the control knobs.
- Governments are very important players in health systems. They can often adjust a control knob both through their role in setting national policy as well as through their role as a direct actor in the health system.

In Chapter 4, we argue that the causal pathways determining performance outcomes are complex. The health system diagnostic tree is useful for highlighting multiple links between causes and the intermediate and core performance criteria. For significant reform, it will almost always be necessary to use more than one control knob to bring about change in performance. Many countries are now considering ways to increase and restructure the financing of the health system, through new social and private insurance schemes for example. Sound design and governance of these schemes inevitably requires new regulation to make insurance effective or to assure access to insurance for those who need it. Behavior change programs may be needed to encourage appropriate use of insurance and discourage abuse by both patients and providers. While we present the control knobs individually, most real reform efforts will require initiatives linking two or more of them.

A related but different point is that action to bring about change through one control knob will often affect the settings on other control knobs. The new resources raised through social insurance will probably affect the mix of payment methods received by providers, altering the overall picture of financial incentives in the health system. Insurance may also channel new funds to some providers (for example, those owned by

the social insurance organization in some countries) increasing their size and importance in the market and thereby altering competition and the mix of providers – settings on the organization control knob. These interactions may sometimes be secondary effects of adjusting one particular knob, but more importantly they can also be intentional – that is, settings may be changed on one control knob in order to change the settings on another.

Our third point is that governments act in different ways to affect control knob settings – they set policy affecting many institutions in the health system, but they are also often large direct actors themselves. This is especially evident in using the organization control knob, because governments often run directly a large part of the health care delivery system through public hospitals, clinics, etc. Changes made by governments in that system, such as development of new health facilities or new systems of management, will, if they matter at all, affect both government and non-government providers through the marketplace. But government can also affect competition without changing how government providers work very much. You can make similar arguments about the other control knobs, where government often plays both a "rowing" and a "steering" role in the health system.

As you read through the next five chapters, each dealing with one control knob, keep these cross-cutting issues in mind. Your challenge as a reformer is to use the control knobs effectively to improve system performance. More often than not, this requires a deft hand on several knobs at the same time.

4. Factors Beyond the Control Knobs

Beyond the control knobs, many other cultural and structural variables also influence the outcomes of the health sector. For example, understanding why a country's regulatory agencies do not function effectively may well require us to understand a nation's political institutions and cultural traditions – factors that may not be easily affected by health reformers. We will emphasize the role of the control knob variables exactly because they are controllable. However, to understand the problems they confront, reformers need to consider other, less controllable variables, like culture, history, and tradition. Engineers designing automobiles can not change gravity, but they have to work with its effects in mind. Below we briefly explore four contextual factors beyond the control knobs, which influence the processes of health reform.

Social and Political Attitudes: A nation's social and political attitudes play a major role in shaping health sector policy. Such attitudes are rooted in a nation's particular history and culture. The egalitarian distribution of health care services in Cuba, compared to most other Caribbean nations, is explained in part by its history since the Cuban Revolution. Such factors determine what kinds of problems are judged important for reform and what kinds of initiatives can be undertaken to address them.

Attitudes Toward and Capacity of the State: Health sector reform possibilities are also heavily influenced by attitudes toward and the capacities of the state in any given

country. In some societies, there is a strong tradition of professionalism in government, and government work is honorable, even prestigious. The state is trusted to handle difficult tasks and it does them reasonably well. At the opposite end of the spectrum, there are nations in which government work is seen as an opportunity to enrich oneself, one's relatives, and one's supporters. In these cases government is often an object of scorn and suspicion. The prospects of health sector reform are of course affected by such factors. Why propose programs that require substantial technical sophistication in government or extensive voluntary cooperation from the business sector, if neither is likely to be available? Why propose greater power for managers or devolution of control over hospitals to local boards if such changes will only lead to increased nepotism or corruption?

Attitudes Toward Health Care: A third cultural feature that bears directly on health sector reform is the attitude of the population toward the health care system and hence how providers inside the system are treated. What is the status of health professionals with regard to education, compensation, and political influence? Cultural norms also shape the financial relationships between doctors and patients. In some societies, doctors feel a moral obligation to care for all those who are sick, regardless of their ability to pay. In other societies, commercial standards apply. In some countries added "informal," "envelope" or "gray" payments are made to public sector doctors. In other states this would never occur. Countries also vary greatly in their stance toward traditional and alternative medical practices. Some societies with extensive traditional medical systems (e.g., India, China) offer substantial support for and utilization of such practitioners. How much should existing norms be accepted, even used, and how much should they be worked against?

5. Conclusion

Now you are ready to dive into the mechanics of health system change. The individual control knobs presented in the subsequent chapters are important tools of reform. But their linkages also need to be understood and managed, as do other important factors which affect their use.

Real change is never easy. This truism may seem obvious but it is often ignored. We have repeatedly encountered situations in which the relevant officials in a country complain bitterly about its health sector and yet veto all suggestions that might actually produce real change. "The doctors will never agree." "The Ministry of Finance would laugh at that idea." "We could never get the unions to accept that," etc.

Given this pattern, we urge readers as they go forward into the detailed discussion that follows to remember one of our initial points. Health sector reform is always political, and is so at all stages of the health reform cycle. That is why we have included Chapter 5. As you consider what you can and cannot attempt on your own country, remember what we said there—turning any of these control knobs is always a matter of political *skill* and political calculation. And bear that in mind in the discussion that follows.

Chapter 7
Financing

1. Introduction

The first of the control knobs we want to consider is financing – the way in which money is mobilized to fund health-sector activities, and how it is used (i.e. allocation.) In this, we will distinguish financing from consideration of how the money, once raised, is paid out. That set of interventions, the payment control knob, will be considered in the next chapter.

Of the five control knobs, financing perhaps is the most important in its impact on the performance of a health system. How a system is financed determines how much money is available, who bears the financial burden and controls the funds, and whether health-expenditure inflation can be managed. These factors, in turn, determine who has access to health care, who is protected against catastrophic medical expenses, and the health status of the population.

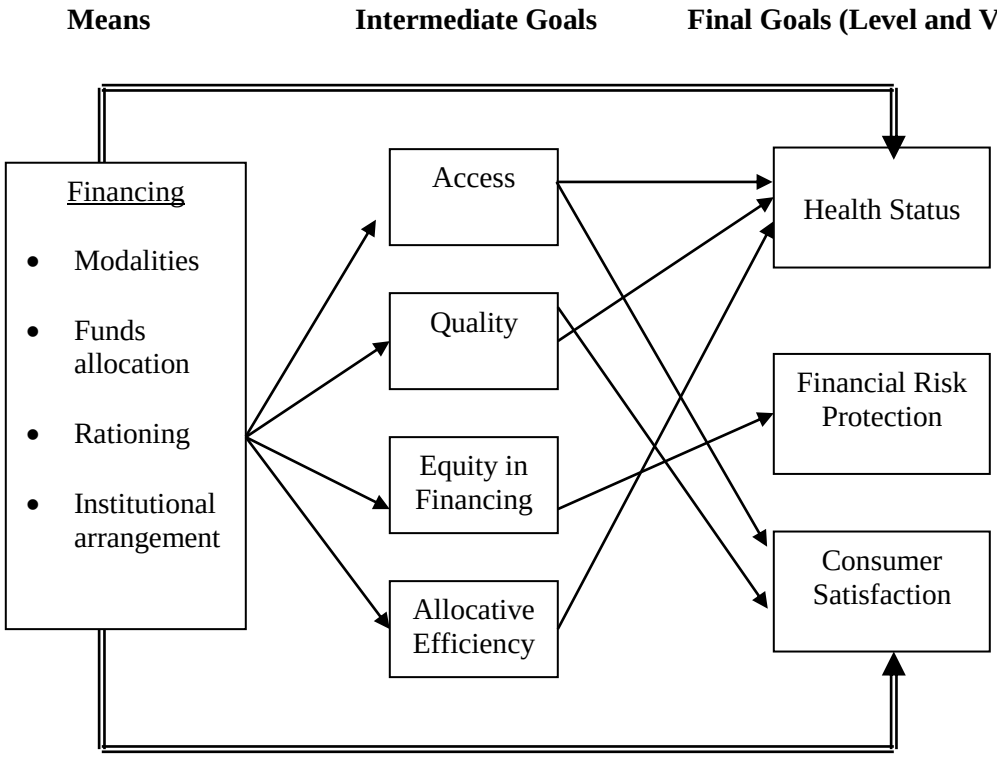
One point should be highlighted: all money raised through any financing modality comes, directly or indirectly, from citizens. Health activities must be paid for. There is no other place for government or private firm, as entities apart from the people, to get resources. Taxes and insurance premiums are paid by citizens – but ultimately, so is everything! The real financial questions are, what amount of funds can a nation mobilize, who ultimately bears the burden, and who receives the benefits.

This chapter is organized into four sections. The first presents the relationship between financing and the goals of a health system. The next section lays out the key criteria that could be used to evaluate the *overall financing policy* when a nation considers which combination of financing modalities may be best suited to it – the mixture depends greatly on social values and politics. The third section describes the different modalities of financing and discusses their advantages and disadvantages. Finally, we discuss the financing schemes that seem to be feasible for countries at different stages of socioeconomic development.

2. Financing and Health System Outcomes

In Chapter 4, we argued that a health system is a means to achieve societal ends. Financing is a major control knob that affects the intermediate and final outcomes. In Figure 7.1., we illustrate how the instruments of financing may influence these outcomes.

Figure 7.1: Relationships Between Financing Instruments and Goals



3. Principal Considerations in Financing Policy

A nation has to decide on a financing strategy – the use of a combination of financing modalities to fund its health system. Five financing modalities are described in section 5 of this paper. They include general revenues, social and private insurance, direct out-of-pocket payments by patients, and community financing. All countries, other than socialist ones, use a combination of modalities. For example, the U.K. is widely perceived as a nation that relies wholly on general revenues to finance its health system. In reality, however, only 76% of its health funds come from general revenues; 12% comes from social insurance contributions, 10% from private insurance, and 2% from direct out-of-pocket payments. When nations search for better financing strategies to improve the performance of their health systems, several essential factors have to be considered.

3.1. Socioeconomic Development and Health Financing

To state the obvious, the same structure in financing and its organizational arrangement cannot be applied to all nations. Health-care systems differ enormously across countries, in particular according to their socioeconomic development – What works in the U.K., may not work in Kenya. On the other hand, do we have to treat every

nation differently? Can't we group nations into somewhat homogenous categories and derive general conclusions for the nations within each group?

A nation's capacity and ability to mobilize funds using different financing modalities are highly correlated with its per capita income. Income determines household capacity to pay for health care and demand for services. Besides household income, several other major factors are also highly correlated with income. These include the tax base from which a nation can raise tax revenues; the number of workers employed in the formal sector, which influences how much can be raised through social insurance; and the number of poor households that have to be subsidized. Lower income nations, on average, have a much higher percentage of their economic activities falling into the category of the shadow economy (unrecorded legal income and proceeds from illicit activities). Friedrich Schneider (1999) estimated that the shadow economies of Mexico, Philippines, Egypt and Thailand are more than fifty percent of their official GDP, while high-income nations such Japan and the U.S. have shadow economies of less than ten percent of their official GDP. It is impossible to collect taxes or social insurance contributions from the income of shadow economy. As for a nation's ability to use any modality effectively, this depends on its infrastructure and competency in public and private management. They too are highly correlated with its per capita income. For these reasons, we use per capita GDP as a first approximate criteria to group nations into somewhat homogenous categories.

We group nations into groups so we can analyze what choices in financing modality might be available, and the amount of funds that can be raised through the various sources. At the descriptive level, we observe that in low-income countries, tax funds usually finance 40-60% of the total health expenditures, 10-15% financed by social insurance (most likely covering civil servants) and 40-50% from the patients' direct out-of-pocket payments. Private insurance is negligible or nonexistent, because few households can afford to buy private insurance and the necessary administrative safeguards are not in place to prevent fraudulent insurance claims. As a country industrializes and its per capita income grows, social insurance usually expands, because the number of workers in the formal sector grows. Private insurance begins to emerge but plays a very small role. The major portion of the total national health expenditures is still financed by tax funds or patients' direct payments.

The major difference between low- and middle-income countries is the relative proportion of total health expenditures financed from different sources, and the size of the private- and social insurance-funded services that offer a noticeable higher quality of services through a compartmentalized health-service market. At the high-income level of development, all nations, except the U.S., established a system of financing to assure universal equal access to health care.

Table 7.1. presents examples of nations in each of the three stages of health development, as well as a broad summary of the financing and organization of health care in that stage. A more detailed description for each stage follows.

Table 7.1. Evolution of Health Care Financing and Provision Systems at Various Stages of Economic Development

	Stage I (three-tiered system)		Stage II (compartmentalized financing and provision)	Stage III (universal coverage**)
	Poor (less than $1,800)*	Low ($1,800-$4,800)*	($5,000-$12,000)*	(greater than $12,000)*
General Revenue Financed + Donor	Public health, prevention Public health services (clinics, hospitals)		Public Health Service	NHS (U.K., N.Z.) Medisave + Cat. (Singapore)
	(50-60%)	(40-50%)	(20-40%)	
Social Insurance	For civil servants only	(10-20%)	Social Insurance (Dir/Indirect Provision)	National HI (Canada, Australia) Bismarckian Social Insurance (Germany, Japan)
			(30-60%)	
Private Insurance	Negligible	(5-10%)	Private Insurance (15-40%)	Managed Care + Medicare (U.S.)
Self-pay	Private hospitals & clinics Pharmacists Indigenous providers		Self-pay	Self-pay
	(35-45%)	(20-40%)	(15-25%)	(15-25%)
	Mali, Nigeria, Tanzania, Kenya, Yemen, Bangladesh, India	China, Egypt, Peru, Ecuador, Philippines, Indonesia	Turkey, Chile, Mexico, Argentina, Brazil, Lebanon, Venezuela, Thailand, Malaysia	* GDP per capita, 1997 PPP $ ** Except U.S. & Hong Kong

3.2. Fiscal Capacity

A key question for a financing strategy is, can it mobilize enough money to meet the desired level of expenditures in the health sector? Fiscal capacity is contextual. The impact of any strategy will depend on the economic structure of a society (e.g., are there many workers in the formal sector), and the government's administrative capacity. Furthermore, "capacity" is really not a yes/no, discontinuous variable; instead, the amount of money that can be raised by a revenue source may well depend on how much a country is willing to give up in terms of other objectives (e.g., in international competitiveness from higher labor costs) in order to increase health-sector spending. Still,

it is useful to think of the match or mismatch between a given financing strategy and a nation's health sector funding goal; this is what is sometimes meant by "sustainability" in such conversations.

The amount that can be raised from general revenues for health depends on a nation's tax base. The amount of general revenue money that a country is willing to spend for health depends on how much a country is willing to give up in terms of other objectives, such as spending for defense or education.

As for social insurance, its capacity largely depends on its ability to collect the contributions from those employers and workers who are covered. Empirical evidence tends to show that social insurance can be effectively implemented only for those workers who are employed in larger companies (e.g. more than 10 employees) in the formal sector.

Private insurance undoubtedly has some capacity to mobilize funds from those who have the capacity to pay and wish to be insured, and who are otherwise unreachable by other financing methods. Upper-income people may be much more willing to buy private insurance covering expanded or higher quality of services for themselves than to pay for general taxes that benefit others. But by the same token, funds mobilized in this way cannot be used to help the poor so easily. Those who do not have private insurance – not because they do not want it, but because they cannot afford it – can end up in desperate straits.

Community financing, as applied to the rural communities of farmers and peasants, has only a limited capacity to raise large sums. At the same time, they can provide a way of mobilizing increased resources in effective ways to produce modest but important gains to fund primary health services and some hospital care.

Finally, direct out-of-pocket payment by patients is a common feature for most low- and middle-income countries. Recent National Health Accounts (NHA) studies suggest that there may be substantial willingness to spend, and ability to spend, on outpatient care, even among relatively poor people in relatively poor countries – especially when we include "informal" payments. The NHA studies show that even in countries with extensive tax-funded public health services, independent private practitioners provide quite a high proportion of outpatient care, paid for directly by patients. In higher income countries, selected services may not be covered by insurance mechanisms (e.g., drugs or dentistry). In other situations, patients may pay something even for publicly provided or insured services (user fees or co-payments); there may also be gaps in what insurance covers in the form of deductibles (at the front end) or limits (at the back end).

3.3. Efficiency in Raising Funds

Raising funds requires spending money to create records, accounting and auditing systems, administrative agency, etc. If the scheme relies on the private sector and competitive markets, money would also be spent for marketing, sales, and for profits.

Thus, an important question is how much do the various modalities spend for administrative expenses and profits. Which ones are more efficient in terms of having larger portions of their revenues raised spent for health care? This concern, of course, is related to the efficiency and cost of health care.

Tax collectability is the major practical concern that leads countries with weak administrative systems to use those taxes that are easiest to collect (e.g., import duties and value-added tax), even if they are not desirable revenue sources on other grounds. Consumption taxes are often easier to collect than income taxes as there are much fewer businesses from which to collect for consumption taxes than there are households from which to collect income taxes, and businesses tend to have better records. Moreover, while income from "black" or "gray" sources can be taxed when it is spent, is not likely to be reported to income tax collectors.

Both payroll and consumption taxes can be difficult to administer when, as is the case in many poor countries, there are many small sellers in the "informal" sector. The reach of any system will, as noted previously, depend greatly on the attitudes of the public toward government in general, and tax paying in particular. Levels of voluntary compliance vary enormously both inter- and intra-nationally, and using schemes that the population sees as legitimate are much more likely to be successful. Again, many of these issues of enforceability will arise again on regulation Chapter 9.

One very serious objection raised about private insurance is in terms of its administrative cost. Drawing on the U.S. experience, insurance companies often spend 25% to 30% of total revenue on expenses other than patient care (sales, administration, and profit). In addition, providers (doctors and hospitals) also have substantially higher administrative costs in the U.S. than they do overseas. Studies suggest that the most advanced American hospitals average two full-time administrative employees for each occupied bed. Competitive insurance systems significantly increase transaction costs. The right regulation (e.g., specifying uniform claim forms to be used by all insurance companies) can somewhat lessen this burden; evertheless, it remains substantial. All of this indicates that private health insurance markets are very complicated systems. Nations who go this route, therefore, need to be prepared to develop substantial capacity to monitor, analyze, and regulate these markets, if they are going to achieve certain kinds of social objectives.

3.4. Political Feasibility

In a democracy, decisions about how to spend money, and about how much to spend, are a quintessential expression of the power of government. From the perspective of democratic political theory, it is important that these choices be made in a way that allows citizens appropriate control over the process. Several features of the organization of the financing system affect this accountability. Is someone who has authority over the financing process themselves subject to selection by democratic political process? Even if the actual administrator is an appointed technocrat, does a chain of accountability connect them to someone whose authority is subject to electoral decisions?

Interestingly, the ideal of political accountability is controversial in quite a few nations. Where distrust for the political process is great, some prefer that health-care financing decisions be made by agencies that are less closely connected to political processes. This tends to be particularly true of interest groups inside the health care system in countries where public budgets are under great pressure. These groups hope that dedicated revenue sources and isolating health care financing from the general budget process will increase the flow of resources both into the health-care sector in general, and into their own pockets in particular.

The public's understanding of, and attitude toward, the "social contract" of a social insurance scheme will have a great impact on its fiscal capacity. Segregating a large revenue pool out of the general fiscal system, and tying it to health care, is attractive to many because the health budget does not have to compete for funds with other national priorities. This feature of social insurance helps explain the enthusiasm for it as a financing option among providers in many middle-income countries. Doctors from Poland to Argentina have been eager to escape the competing pressures on government tax revenues by creating their own dedicated funding sources. Whether this is in a nation's best interest is, in turn, a very complex political judgment whose dimensions are beyond the limited scope of our current analysis.

4. Key Considerations in Selecting Financing Modalities

Several key factors have to be considered in deciding on the financing modality and its organizational arrangements,

*Equity***:** Since financing directly affects the distribution of the cost of the health care, the most obvious question is who bears the financial burden. Since how the funds are used directly affects the distribution of health care, the obvious question is, who receives the benefits. The criterion to use for assessing the distribution of burden and benefit is equity, which, as we discussed earlier, has several dimensions: v*ertical equity* refers to the distribution of the burden between the rich and the poor; h*orizontal equity* has to do with creating fairness among those at the same income level, including those live across different regions.

*Economic Effects***:** When government compels a firm or individual to pay a tax on a selected activity such as wages, the firm and individual may alter their decisions as to how much to work and earn. Various options do have differential impacts on deterring or encouraging investment, employment opportunities and labor supply, thereby influencing economic activity in the short and long run. What are the financing modality's implications for *macroeconomic growth and efficiency*? We, of course, want to limit the negative macroeconomic impacts.

*Implementability***:** Another aspect of judging any financing options has to do with whether or not the scheme can be effectively put into practice in a particular national context. This, in turn, depends on two key features of the local situation. *Social acceptability:* government's capacity to coerce citizens is, in fact, quite limited.

Obligations that governments impose work best if citizens exhibit high levels of voluntary compliance. For financing schemes, the question is, Will most of those who have to pay will comply, or will they cheat? *Administrative Capacity:* does the nation have the organizations, management skills, and human resources to operationalize a financing scheme effectively?

5. Options in Financing

Our typology of financing options is a mixture of the fiscal and organizational aspects of the money-raising process. It reflects both the economic sources used by various alternative modality, and the organizational arrangements through which financing is carried out. Our options are best thought of as a set of "ideal modalities" – each alternative is a highly stylized simple "pure case."

However, most actual financing systems are *not* pure. Instead, most nations have chosen a "mix and match" approach, combining various ideal modalities in complex ways that reflect the trade-offs in objectives that a nation is willing to make, based on its particular national, political, and economic history. For example, general revenue-based systems often require patients to pay user fees, and do not cover certain kinds of medical expenses (e.g., drugs). Likewise, most private insurance schemes require patients to make substantial direct payments to providers.

Our categories are not meant normatively (to say what should be done), but descriptively; that is, to provide a framework for characterizing and analyzing various complex actual cases. We discuss the five major modalities of financing below.

5.1. General Revenue Financing

In general revenue financing, the entire menu of tax revenue sources is being used to support the entire range of government activity. In this scenario, t*he health system has to compete for funds* with other government-funded programs and gets its resources via the regular government budget process.

Government obtains its general revenue from taxes and fees. The usual taxes are income taxes, corporate profit taxes, value-added taxes, grain tax, and excise taxes. Each one places a tax burden on different groups of citizens. The mix of tax revenues for a particular country depends on its economic development and on people's confidence in the government, which shapes their willingness to pay taxes just as it does government's ability to collect taxes.

Some tax systems, such as progressive income taxes, and taxes on profits, are more sensitive to both short-run business cycles and long-run growth than others. Growth moves people into higher tax brackets, increasing marginal and average collections for income taxes, while profits remain sensitive to the business cycle, whose sales go up in good times. In percentage terms, taxes on luxury goods also move most

from this perspective. Payroll and sales taxes, in contrast, move slightly less than Gross Domestic Product (GDP).

General tax financing offers greater political accountability in democratic political systems, since the key decisions do pass directly through the legislative process. Thus, the level of funding can be controlled and decision-makers can be held accountable for their actions. On the other hand, implicit tax support through mechanisms like tax concessions (e.g., allowing expenditures on health care to lower a person's income tax liability) is neither transparent nor controllable. It is difficult to know how much revenue is lost via such schemes, and difficult to control who benefits, and for what purposes.

The amount of money raised by broad-based taxes does tend to track the level of prosperity in a county, in both the long- and short-run. Such a financing system will therefore tend to automatically generate more resources for health as the economy grows over time. On the other hand, collections also rise and fall in the short-run over the business cycle; thus, tax financed health care systems are likely to be especially hard pressed in a recession when tax collections fall, and the competitive demand for other kinds of social spending rises.

One key issue that has to be addressed in general revenue financing for health care pertains to which level of government should pay for it. If health services are jointly financed by the national and regional governments, how will revenue be transferred among governments? What activities will the national taxes support, and will they be tied or targeted in some way?

5.1.1. Vertical and Horizontal Equity of Alternative Taxes

Who bears the burden of tax funds used for health programs? Economics tells us that the tax burden may not fall simply on those who pay taxes: higher health costs may not force tax rates up; they may instead, lead government to curtail other services when it appropriates more money for health. In this case, the potential beneficiaries of other services (those that are now *not* being provided) are the ones paying for higher health sector costs. The same is true when government subsidizes some activities through tax concessions. The payers in that case are one of two groups: they might be those who pay the other taxes raised to make up the revenue lost through such concessions, or they might be the potential beneficiaries of other activities who now are no longer served because tax concessions have made the money for those other services unavailable. The extent to which government raises taxes or curtails other services determines which groups are the real bearers of the financing burden.

In addition, those who directly pay any given tax may not be those who ultimately bear the burden – employers shift some of the cost of payroll taxes back to workers in the form of lower salaries. Similarly, property taxes on landlords may be passed on to renters in the form of more expensive rents. Whether such shifting occurs, or to what extent, depends on a variety of features of the relevant markets: How will the number of employees drop if employers offer lower overall compensation? How easy is it for

employers to substitute machinery for labor? Will buyers buy significantly less, if prices are raised? There is vast economic literature on such issues.

Vertical Equity: In an earlier chapter, we introduced the distinction among progressive, proportional, and regressive taxes as a way to judge the vertical equity of financing systems. These distinctions are determined by the percentage of income taken from those at various income levels. It is worth noting that only comprehensive income taxes, including taxes on capital income, can be truly progressive in most countries – only such taxes can take a higher proportion of the income of those at higher income levels. In contrast, payroll taxes are mildly regressive, because the proportion of total income that comes from payrolls falls at higher income levels; payroll tax payments therefore fall as a percentage of income, as income increases.

Indeed, if there is a limit on any one person's total contribution (as is sometimes the case), the scheme becomes even more regressive. Taxes on consumption – sales and value-added taxes – are also moderately regressive, because the percentage of income that is spent and taxed, as opposed to saved, is lower at higher incomes. This is why some consumption tax schemes exempt, or tax at a lower rate, commodities on which the poor spend a higher percentage of their income, like food. For similar reasons, we often find a higher tax rate on luxury items where the rich spend more of their income.

Similarly, the vertical equity of how the funds were spent can be assessed by examining the distribution of benefits by income groups. In the final analysis, we are interested in the vertical equity of net benefits; that is, the benefits minus burden.

Horizontal Equity: The horizontal equity of tax-based financing depends on the kind of taxes. Broad-based national taxes do not raise horizontal equity issues, but narrower taxes (e.g., on tobacco or alcohol) may do so. Is it really fair to ask drinkers or smokers to pay more, especially if, as appears to be the case, moderate alcohol consumption improves health?

In recent years many countries have shifted both control and fiscal responsibility for health care to the provincial or local level. Such schemes can raise serious horizontal equity issues since more prosperous areas can finance the same service as poor areas at lower tax rates, or better services with the same tax rate.

For this reason, nations that have gone the route of fiscal decentralization have often found it necessary to establish interregional equalization funds. The countries with the longest standing systems of this type, like the Scandinavian countries and Columbia, now rely heavily on national taxes to supplement local sources.

Constructing an equalization scheme can be quite complex since it has to do more than just take account of need (population, illness, and income). It also has to create incentives to prevent poor regions from lowering their own tax effort and "free riding" on the collective efforts of their neighbors. Decentralizing taxing decisions also runs the risk of creating a "race to the bottom," as regions compete in attracting business by lowering

taxes. Richer (and more politically powerful) regimes may also object to the creation of an interregional system of redistribution, precisely because it does threaten to eliminate their own advantages.

There are also administrative issues as we use ever-smaller areas as a tax base. The question of where income is earned or spent (and therefore who gets the tax proceeds) becomes steadily more complex. This is one reason that local governments often tax non-moveable assets like land: at least they know where it is!

5.1.2. Economic Effects

The economic effects of using general tax revenue to finance health care will depend on what specific taxes are employed. In theory, very high income tax rate should lower upper income individuals' willingness to work; in fact, the limited empirical evidence does not support this concern. Studies tend to find that for upper-income individuals, their labor supply is insensitive to marginal income tax rates. High labor costs can also decrease a country's international competitiveness, and/or slow the pace of investment and economic growth. As investment capital becomes increasingly international, the effect of such considerations are liable to increase.

We note here a somewhat obscure issue about taxes – an issue of great concern to some economists – the concept of *"excess burden."* When taxes are imposed on goods, buyers pay a higher price than they would otherwise, and sellers gets less – the tax drives a wedge between what is paid and what is received. As a result, less is bought and sold than otherwise would be in a perfectly competitive market with zero tax. From an economics perspective, this loss of output poses a problem, because customers would have been willing to pay more than the cost of production of the goods that are no longer produced. Yet, thanks to the tax, they no longer get the benefit of this gain. And the larger the adjustment in output due to the tax, the greater the loss will be, as measured by the difference between the willingness-to-pay of customers and the cost of producing the lost marginal output. This difference is what is called the "excess burden" of a tax.

In response, some economists argue for taxes that lead to the smallest distortion in outputs. This turns out to imply higher taxes on to goods where customers care so much, that they continue to buy nearly the same amount, even as prices rise. Economists call such demand "inelastic." Ironically, such goods are also often the ones that people purchase because they are very important to them – like medical care and basic food. Taxing goods whose demand is unelastic is thus likely to violate our concern with vertical equity, which does not stop some economists from advocating such an approach.

5.1.3. Summary

General revenue is a major source of financing for health care in rich and poor countries alike. International experience tends to show that as a country becomes more affluent, its tax base becomes larger and the government's ability to collect tax becomes stronger, a larger share of the health expenditures is funded out of general revenues. General revenue financing is also a politically controllable and accountable source of

revenue. It can be, but does not have to be, tailored to meet both horizontal and vertical equity concerns, and has good long-run growth potential.

On the other hand, its very controllability makes it unattractive to some – especially when the political system is weak and subject to endemic corruption and favoritism. When government functions badly, there may be problems with public acceptability, hence collectability of tax revenue – especially in nations with lesser administrative and enforcement capacity and/or an especially cynical or tax-avoiding citizenry.

Certain kinds of services (public health and preventive services, and/or care for the poor) are often tax-supported, even when other methods dominate a nation's health care financing system. This is because taxes can be redistributive, and can be raised from a broad base. In countries with well-functioning economies and strong administration, tax financing can generate substantial funds for the health sector. But these conditions are not met in all countries, which helps explain the widespread interest in our second financing option – social insurance.

5.2. Social Insurance

Three characteristics distinguish social insurance from private. First, social insurance is *compulsory* – everyone in the eligible group must enroll and pay the specified premium (contribution). This contribution is most often specified as a percent of the wage – the economic literature calls it a "dedicated payroll tax". Once a person has paid the minimum number of payments he or she is entitled to the specified benefits.

Second, social insurance premiums and benefits are described in social compacts (laws) established through legislation. The contribution rate is not easily adjustable by merely administrative action; instead, it is specified in law or in a difficult to change regulation. All this creates an implicit bargain or social contract between those covered and the system. Citizens agree to pay a certain amount with some confidence that it will be used fairly and effectively, to reliably fund health care for all who are part of the system.

Lastly, while most ministers of finance object to earmarked payments for social insurance programs, they do have the advantage that the insurance benefits demanded by the voters come with a clear price tag – they have to put their money where their mouths are.

Social insurance schemes do not necessarily cover everyone; most schemes were designed just to cover the workers in the formal sector. When a country desires to provide universal coverage, the government often has to use general revenue to subsidize the contribution to be paid by pensioners and the poor. Costa Rico was the first nation to achieve universal coverage in 1941 by having general revenue pay for the farmers and low-income and poor households.

The organizational form for social insurance consists of two types. Under the Bismarkian model, plans are established and managed by many non-profit organizations, often divided along industrial, geographic, and occupational lines. They are commonly called sickness funds, and may compete for enrollees under some strict regulations. Most social insurance plans in Europe and Latin America are this type. The multiple organizational arrangement allows for some variation and choice in the benefit packages; however, it suffers from adverse selection and high administrative costs. To remedy in part these shortcomings, several nations, including Canada and Taiwan, have the program administered by a single parastatel agency that is an organizationally distinct entity, outside of the detailed day-to-day control of the political process.[18]

5.2.1. Equity in Financial Burden and Benefit

Given the potential variety in detailed specifications, it is not surprising that the equity effects of actual social insurance schemes depend on their details. In any given society, the vertical equity of the payroll tax system will depend on how much income of the rich escapes the tax net because it is from capital, not wages. Horizontal equity will depend on whether some groups (e.g., small businesses, farmers) escape paying fair tax burdens due to legislative favoritism or simple evasion. Horizontal equity will also depend on whether there is really only one plan; if there are many, then some plans (because of a more favorable mix of members) can offer their cohorts cheaper or better care for the same price.

This last point raises a structural issue about all non-universal social insurance schemes. Those with upper income workers can provide better care at the same tax rate and are also likely to have less sick populations, given the correlation of economic and health status. The poor tend to be sicker and the sick poorer than average.

5.2.2. Economic Effects

Economic theory and observation both lead us to believe that over the medium-run of a few years, workers will pay for the largest share of health insurance premiums (either directly, or in the form of lower wages) – even where employers nominally contribute a share on their own. Employers have a certain willingness-to-pay for various kinds of workers. At some level, they may not care whether this is in the form of wages or fringe benefits. When the employer payment of premium results in lower wages,

[18] Economic literature often describes social insurance financing as indistinguishable from government tax financing. Social insurance experts vehemently disagree (Robert Ball, Robert J. Meyers, 1994). Economists view social insurance as a tax-financed program mainly because the contribution is compulsory. However, economic analyses ignore many major social and institutional differences between a general tax revenue financed insurance (e.g. Canada) from social insurance (e.g. Germany, Japan, Taiwan). The differences are: (a) social insurance is not a right of all citizens but only covers those who are eligible and have met the minimum contribution requirements. (b) The benefits are usually related to the contribution base. (c) People perceive that they have paid a premium contribution in exchange for the right to specified benefits. In other words, the benefits are not welfare from the government. (d) Contributions (premiums) paid for social insurance programs are earmarked for those programs and are separated from general taxes. Social insurance is required to maintain its own solvency, and thus has greater transparency and accountability to the people. Contribution rates and benefits cannot be unilaterally changed by executive government decision. The social compact can only be altered through new legislation which requires consensus and the support of all interested parties.

economists term it as the premium cost being shifted backward-- to the workers. The ability of employers to shift the cost to workers rather than to higher prices depends on the conditions in the labor market, including how strong the labor unions are.

The question is, What does this do to economic growth? If total labor costs are really not effected by premium changes, then the answer should be – not much. The real impact is hard to assess, because real labor markets are full of all kinds of rigidities and imperfections. In a situation when two-thirds of employers' premium payment is passed back to the workers, one-third will ultimately be paid by customers via higher prices. This could injure a nation's capacity to compete.

5.2.3. Implementability

Implementability is often one of the great attractions of social insurance schemes. In many industrial economies, large enterprises employ a significant share of the workforce, and even small businesses have reasonable records. Moreover, the "social contract" can increase worker and management willingness to pay, as opposed to evade a non-dedicated payroll taxes, because a dedicated fund, with distinct administration, is more trusted to deliver something of value to members. Citizens think, "At least those dishonest politicians cannot get their hands on 'my' money."

On the other hand, such hopes are not always rewarded. The social insurance system can be seen as just another government function and be associated with past illegitimate regimes. As a result, in some postcommunist countries, for example, tax evasion is widespread in the social insurance system. Thus, Hungarian small business groups have used their position on the managing board of the social insurance fund to ensure favorable treatment of their own members. In response, large businesses have begun to shift as much employee compensation as possible from cash, which is taxed, to in-kind support (e.g., cars, houses) which is not. Add in a government reluctant to fund its "share" for pensioners, etc., and you have a system without either the fiscal capacity or the popular acceptance some of its advocates hoped for.

5.3. Private Insurance

Private insurance is distinguished by the buyer voluntarily purchase insurance from independent, competitive sellers (either for-profit, or non-profit) who charge premiums that reflect buyer's risks rather than their ability to pay. Voluntary purchase of insurance can be made on an individual or group basis.

Around the world, there has been in recent years increased interest in various forms of private insurance as a mechanism for health-sector finance, which appears to be driven by several different lines of argument. The first idea is that private insurance will mobilize additional resources. For instance, since non-payers do not get coverage, the problems of tax evasion can be avoided. Advocates of private insurance also argue that when people can choose a plan and a carrier, they will feel more empowered and will become more willing to pay for health care.

A second contention is that those with different attitudes and values, including those at different income levels, will prefer different health insurance plans. It is claimed that a competitive market will respond by offering a differentiated range of products, something a publicly controlled social insurance monopoly would have neither the incentive nor the inclination to do.

5.3.1. Equity in Financial Burden and Benefits

How does private insurance stack up from an equity perspective? The answer is mixed at best. Premiums based on individual or group risks are extremely regressive. Indeed, given the inverse relationship between economic and health status, true risk-based premiums lead the poor to pay more. Even regulated uniform rates that do not vary at all with income are more regressive than almost any other funding option. Similarly, from a risk protection viewpoint, competitive private insurance markets leave a lot to be desired, exactly because low-income individuals and families cannot afford to buy insurance – the ones most at risk of serious financial costs of illness are those least able to afford it.

From the viewpoint of horizontal equity, the question is, what dimensions of comparison do we care about? Private insurance markets treat those with similar risks, or in similar risk groups, with the same cost—those in high-risk groups pay more. It is a matter of philosophy and policy as to what differences we want and do not want, to affect individuals' health care costs.

5.3.2. Economic Effects

One argument for private insurance relies on competition to improve efficiency. The claim is that competitive insurance markets will lower health care costs: insurers who are eager for customers will cut prices, and to make money at such rates, they will effectively pressure providers to cut what they charge. Providers faced with lower revenues will have every incentive to reorganize their work in order to reduce cost. The collapse of communism, the prestige of market ideology, the crusading advocacy of some well-funded zealots, have all reinforced the trend toward pro-market perspectives.

The most worrisome failure of competitive private insurance is the question of "risk selection." In a world where more and more disease is chronic disease, health care costs are increasingly predictable on a year-to-year basic – Those who are sick this year are far more likely than average to be sick next year. The 5% or 10% of the people in any insurance pool who are the sickest often account for 60% to 70% of the total cost. Together, these facts create enormous incentives for competitive insurance companies to sell health insurance only to healthy people – Or, if they do sell to the sick, to charge rates high enough to yield a profit, even from such poor risks.

The latter practice, of course, only gives the sick reason to lie to insurance companies about their health. If regulators require firms to sell insurance to all at the same rate, healthy customers are likely to not buy insurance, because it will be unattractively expensive when their lower risk of disease makes the coverage not worth it. This can lead to what is called a "underwriting death spiral": as good risks leave an

insurance pool, losses mount, rates have to rise, and even more of the healthy drop out, leaving expensive, unhealthy customers whose high cost of care demands ever higher rates.

5.3.3. Implementability

In creating a private insurance financing system, nations confront a series of detailed design issues. The first issue is whether contracts will be sold on an individual or group basis. The U.S. market is heavily based on employer groups. This is done in order to limit the incentives for, and possibility of, individuals opting into, or out of, various risk pools and thereby destabilizing the system. Even where groups are small, there has been experimentation with various kinds of collective purchasing to at least lower transaction costs.

The second design issue is the nature of the insurance entities. Are these for-profit or not-for-profit? If they are not-for-profit, are they state enterprises, independent NGO-type entities, some kind of union or community-sponsored cooperative, or some other form?

A third set of choices deals with government's role. What form of regulation, licensing, etc., will sellers be subject to? For example, will governments insist on certain levels of financial reserves, regulate premiums, or set terms for minimum coverage?

A final set of design questions arises as to the relationship of insurance companies to health care providers. Does government exercise any regulatory control over the relationship between insurers and providers? For example, can providers and insurers be combined in some sort of arrangement like an American HMO? Are such options forbidden, required, or optional? Will government regulate how insurance plans pay providers, as the German government regulates the relationship of the sickness funds to the regional physicians associations with whom they contract?

Beyond all of this, there is the question of the role of government in shaping the competitive structure in these markets. How many firms do they allow and/or encourage? What geographic areas or population groups can each competitor appeal to, or operate within?

What should be clear at this point is that the private insurance option does not get the government out of the business of operating the system of health care financing. Rather, it poses a complex set of regulatory and management issues to government that may be quite new and different from those it confronted if it previously relied on general tax financing.

As we suggested above, private insurance markets lead to significant and subtle new responsibility for governments, which they may not be well equipped to undertake. Does government have the capacity for underwriting and financial analysis to set reasonable reserve requirements? Does it know enough about the dynamics of competitive insurance markets – a phenomena which may be new to the country when

companies try to attract the good risks, to create and control such markets effectively? Can it recognize and regulate "cream-skimming" behavior? Will it have to control prices and/or the terms on which policies are sold? Experience with insurance regulation in the U.S. at both the state and federal levels suggest that high levels of analytical competence and political integrity are required to perform such tasks successfully.

The quantity of insurance purchased remains a decentralized decision in the hands of individual businesses and consumers. The declining rate of private insurance coverage in the U.S. demonstrates that as premiums rise, many consumers stop purchasing insurance. There are limits on what government can do about this. Of course, government can offer subsidies to encourage private insurance purchases such as in Chile. But that really amounts to moving away from insurance to tax-based financing by giving a subsidy. Controlling insurance rates has its own difficulties. Set rates too low, and companies will flee the business, as many U.S. insurers have done with regard to offering special policies to retired citizens. Set them too high and you risk making insurance unaffordable for many.

Many advocates of private insurance inside the health sector believe it is a solution to their financing problems. The lure of the market and the enormous sums generated by private insurance firms in the U.S. makes allies of an unlikely array of characters ranging from greedy entrepreneurs to economic ideologists, to cash-starved physicians. But the very high administrative costs and poor equity performance here are not trivial problems with this approach. The need for sophisticated regulations that exercise constant vigilance may be a requirement many countries that are toying with such schemes cannot realistically meet.

5.4. Out-of-Pocket Payments

Individual patients directly pay providers out of their own pockets for goods and services received. These expenditures are not reimbursable by third parties, such as insurance plans. User fees are a sub-category of out-of-pocket payment, and refers to the out-of-pocket payment for public-sector-provided services. Government facilities charge patients user fees to finance a portion of their operating expenses.

Health sector reformers have been interested in such financing approaches for two reasons. First, especially in lower income countries, such payments are seen as an administratively feasible way to raise additional revenue for institutions and activities at the periphery. The idea is that money collected locally will be spent locally. This, it is believed, will diminish "leakage" from graft, corruption, and overhead expenses, as local collection and disbursement increase accountability and transparency. As a result, it is hoped that such arrangements will induce patients to be more willing to pay for care – since doing so will produce evident local benefits.

The second argument is more likely to be made by economists, who fear that giving away health services for free only encourages allocatively inefficient overuse. Their analysis rests on the subjective utilitarian premise that the purpose of the health care system should be to maximize customer satisfaction, as measured by willingness-to-

pay. When services have zero price, economists argue, customers will use these services even when the value to them is less than the cost of the production. As a result, more total customer satisfaction could be provided if those low-valued services were not produced, as the resources that are used to produce them could then be used to produce something of more value to customers elsewhere in the economy. User fees and co-payments are therefore seen as desirable because they avoid the worst misallocation by discouraging customers from consuming those services with the lowest value to them.

Notice that these two arguments are not fully consistent. Those who favor user fees for revenue-raising purposes want to put prices on highly valued services, for their utilization will not change much when those services are charged for. Such advocates may also be relatively responsive to the notion of finding a way to help those at low income pay for such fees; otherwise, their consumption might be heavily affected.

Advocates of pricing-for-efficiency, in contrast, most want to raise prices for services that are both expensive to produce and of low value to customers. For them, declines in usage, as a result of instituting direct payments from patients to providers, are not an objection to their scheme. Rather, such declines are a sign of the appropriateness of the decision to institute prices in the first place.

There is another complexity. Given the possibility of supplier-induced demand, there is no guarantee that all health care services really are of value to those doing the purchasing. If doctors influence patients to consume inappropriately low-value services, allocative inefficiency will persist. Indeed, in practice, whether allocational efficiency increases or decreases when fees are imposed will depend on two offsetting effects. One is the impact of fees on discouraging patients from purchasing services of low value. The other is any incentive effects such fees have on providers to cause an increase in inappropriate use. (The magnitude of the latter effect will obviously depend on a variety of factors including professional norms of the relevant providers, their training, and the ways in which fee income flows to clinical decision-makers.)

The total amount of money raised through direct patient payments is often both unknown and difficult to manipulate. The cost of acquiring good data can be quite substantial, since sophisticated household health-expenditure surveys may be the only way to do so.

In discussing direct patient payment, we cannot avoid mentioning the widespread existence of "informal" or "unofficial" payments in state-run systems, where they are typically implicitly or explicitly illegal. These often arise when budget support for the health care system is diminished by difficult macroeconomic circumstances and providers are desperate to preserve their incomes. Local traditions also matter. In some countries, such fees are expected to ensure good care, not only by doctors, but by nurses as well. They tend to be higher for those providers seen as most "active" or "heroic" – surgery and Ob-Gyn. These payments go under different names in different countries – "gratuities," "under-the-table payments," "black money," "gray money," etc. From an economic point of view, these are just another manifestation of direct payment financing,

as they have much the same effect on demand and the distribution of the burden of support for the system that legal user fees would have.

Recent National Health Accounts (NHA) studies do suggest that there may be substantial willingness to spend, and ability to spend, on outpatient care, even among relatively poor people, in relatively poor countries – especially when we include "informal" payments. That, in turn, raises the possibility of creating mixed systems where fees play a non-trivial role in outpatient settings. On the other hand, it is quite possible that fees set to discourage low-valued uses will also lead customers to curtail use beyond the point that is optimal in terms of health maximization. This suggests the need to find ways to encourage utilization of public health and prevention services. Unfortunately, in a world of chronic disease, many such encounters are not distinguishable from ordinary clinical care, since the prevention occurs *after* some disease has developed (so-called secondary prevention, as in getting someone who has had a heart attack to quit smoking).

Direct payments are often seen as a useful supplement to tax funds as a source of institutional support, but it is doubtful that they can serve as the major fiscal basis for modern inpatient care for all but the very rich. Such care is so expensive, and the demand for it sufficiently infrequent, that the pressure to develop risk-spreading financing (i.e., some form of insurance) is almost irresistible once a country has reached a certain level of political and economic development.

5.4.1. Equity in Financial Burden and Benefits

From both risk protection and equity perspectives, out-of-pocket payment is the worst possible system. Those who are both sick and not very well off financially face the risk of either untreated disease, or impoverishment, or some combination thereof.

From a vertical-equity viewpoint, direct payments are extremely regressive, especially given the correlation of poor health and low income. They are, if anything, even worse than private insurance, which at least offers some averaging possibility when there is group purchasing, or when rates are regulated.

In this context, we do need to note that in many countries, practitioners supported by fee-for-service engage in some forms of price discrimination – charging the rich more than the poor. Yet, while such practices advance equity, they also represent the best way for sellers to maximize their incomes. By charging the rich more and the poor less, the poor purchase more total services, and generate more income, than they would otherwise. This is a net gain to any seller who has access time available and whose marginal cost of providing additional service is less than what they charge the poor. Whether such practices also violate norms of horizontal equity, by being unfair to the rich, is not a question that has received much attention.

There is a particular method of direct payment, *medical savings accounts*, is both sufficiently distinct, and of sufficient interest, to deserve special mentioning. The prototypical example can be found in Singapore. The scheme here was a conscious

strategy to build up reserves in light of an anticipated increase in the elderly, high-high-care-using population. The compulsory savings scheme (based on a percentage of wages) goes into dedicated individual savings accounts, where it can then be used for inpatient services.

Several features of the Singapore arrangements deserve comment. The system is first and foremost just what it says, a savings scheme designed to compel workers to save for large medical expenses and shift current income into the future. Until substantial balances build up in individual accounts, and while patients are still young, they are actively discouraged from using their accounts to pay for care. Instead, great effort is made to encourage them, or their families, to pay out-of-pocket.

Second, to make it possible for people to pay out-of-pocket yet afford basic hospital services, hospitals have multiple levels of service, from "C" class, multiple bed wards, to "A" class single bed room. The lowest class of service is heavily subsidized by tax revenue, making it available to many with modest income (medical services are not supposed to vary, only amentities). However, Singapore discovered that most workers cannot save adequate amounts to pay for inpatient services. In 1990, it therefore revised its savings scheme by adding a catastrophic insurance plan, in which workers pay the premiums by withdrawals from their medical savings account.

Third, to make current payment possible, there is also a system of case-by-case subsidies available to those cases in which social workers determine that neither individuals nor their families can pay what is required.

This complex scheme has been modified and evolved over a decade in order to deal with the otherwise objectionable vertical equity effects of direct payment. Singapore's relatively unique socio-cultural situation is widely acknowledged to play a major role in the scheme's functioning. It is a geographically small, relatively wealthy, largely Chinese society, in which both thrift and family responsibility are deeply engrained values. It also has a competent and sophisticated government fiscal system.

5.4.2. Implementability

The greatest advantage of user payments is, in fact, in the realm of implementability. Because it gives the opportunity for sellers to refuse to serve non-payers, it forces buyers to actually pay. Since transactions occur "on the spot," they can do so using ordinary commercial processes (e.g., checks, cash, credit cards) that do not generate any added record-keeping burden. From the perspective of government, the payment system is essentially self-implementing and does not take substantial public budgeting resources to operate. When user fees are very high, and sellers cannot refuse to serve patients, as is the case for American hospitals treating someone who is uninsured, the costs of collecting the fee after the fact can be substantial.

The incentives for overuse that fees create can lead to substantial countervailing government regulatory efforts, whose cost is properly considered part of the cost of the direct payment scheme that creates the need for such efforts. Many quality initiatives

(like regulatory and licensing providers), as well as price regulatory activities, have developed as part and parcel of direct payment financing systems. So too have some efforts to limit the number of providers to license institutions to constrain potential overuse.

5.4.3. Summary

Patient payments are easy to administer and a potentially effective source of revenue. However, they raise serious ethical questions from the perspective of vertical equity and risk protection. They can also foster inappropriate utilization through their incentive effects on physicians and other providers, even while they discourage inappropriate overuse from the viewpoint of consumers. Nor do they have the capacity to finance universal coverage of expensive services. However, patient payments clearly have been attractive to many countries, because of their power to mobilize resources otherwise unavailable to the health sector.

5.5. Community Financing

Communities organize and control the direct provision of their own primary care services based on (local) universal pre-payment. Under most community financing schemes, the financing and delivery of primary care are integrated, but the financing and provision of secondary and tertiary services are separated. Providers are usually private for-profit or non-profit firms or NGOs characteristic of a full community-financing scheme. As we discuss at the end of the chapter, this complex variety of real-world financing choices has to be judged in light of the particular economic, social, and political circumstances, in each.

Many of the world's lowest income countries, with substantial rural populations, have despaired of finding ways to reliably finance and deliver health services at the village level. Where a ministry of health directly operates clinics in such areas, it is often difficult to get physicians to staff them: they often simply evade or refuse, and/or do not attend regularly, and/or provide poor customer service that is culturally insensitive. At the same time, village residents often have little confidence in those services. As a result, they make extensive use of traditional healers and folk medicine practitioners for outpatient care, and when acutely ill, they flood into, and overcrowd, regional and international hospitals.

It is not uncommon for those same regional and national centers (which are often also teaching hospitals) to absorb a very large share of the nation's overall health budget. This is, in part, because a country's economic and political elite wants to ensure that there are at least a few "world class" institutions in the country in which they can get care. Also, the most prestigious institutions often have substantial political connections and influence which allows them to effectively defend their interests.

In many poor countries, it is difficult to raise added revenue from general taxation. Corruption and/or inefficiency may be widespread, record-keeping poor, informal activity widespread, and tax evasion a cultural norm. In response, there has been increased interest in recent years in a financing scheme that has come to be called

"community financing." The essential idea is to raise and spend money locally, at the village level, for primary care. The theory is that local control of expenditures will produce transparency and accountability. This will, in turn, help ensure honest, efficient and culturally competent services.

Such sources and administrative arrangements will be attractive and credible to local people. This will, in turn, increase their willingness to contribute financially to support these services. Advocates of this approach note the non-trivial sums spent by relatively poor local people on traditional healers, folk, and western medicine. They are intrigued by the possibility of getting access to those funds to support more "mainline" public health and primary care efforts.

In the ideal, typical community-financing scheme, there is a combination of local political accountability, community-operated primary care, and universal pre-payment. In effect, the image is of a community-based, mini-health maintenance organization, organized on the so-called model with doctors paid on salary. Secondary (hospital) care is outside the scope of the ideal-type scheme: it is too expensive, ambitious, and requires too large a catchment area for the maintenance of effective local control. Universal (i.e., compulsory) membership is included, however, in order to get around free-riding by the well, and adverse selection by the sick. Community resources may also be mobilized in other ways – labor to construct buildings, for example.

Many actual experiments in raising local funds to finance primary health care only partially partake of the ideal. It is not uncommon for membership (and payment) to be voluntary, the range of services covered can be quite limited (e.g., a locally financed, revolving drug fund); or the financing can be heavily fee-for-service rather than pre-payment, or come from national tax sources, not local contributions.

There are examples in the world of effective rural outreach with centralized tax financing and service organization. Sri Lanka, Cuba, and Kevoela (in Southern India) all have well-above average health status relative to their income levels and relatively dense networks of public providers, even in rural areas. Advocates of community financing respond that such examples only show the exceptional historical, political, and cultural circumstances that are required to be successful via the centralized state route.

One of the strongest features of the community financing model is accountibility. But, as we noted previously, real communities can be complex political systems characterized as much by division as by consensus.

One of the limits of a narrow local fiscal base is that it can be difficult to sustain in a downturn. Any given rural area is likely to have a less diverse economy than the nation as a whole. Malaysia, for example, has relied, in succession, on tin, tea, palm oil, petroleum, and high-tech manufacturing for its growth over the last thirty years. Commodity price cycles that have seriously injured some parts of the country have been offset by growth elsewhere. An individual village, however, does not have the benefit of the same kind of "portfolio diversification."

5.5.1. Equity in Financial Burden and Benefits

Where community financing is both prepaid and compulsory, it does offer a certain amount of risk protection. Do note, however, the omission of secondary care means that this form of community financing has to be supplemented by other schemes (e.g., tax-supported hospitals) if citizens are to be protected from the financial risks of serious illness. Non-compulsory schemes, where there is adverse selection by the sick into the covered pool, will offer even less risk protection. Given the relatively small and geographically concentrated nature of the populations covered by such systems, they do not have the capacity to do enough risk spreading to provide insurance against adverse events like localized epidemics or natural disasters.

From a vertical equity view, one critical question is the nature of the pre-payment system. In particular, does it have any ability-to-pay component, or variability with income? Per-capita or per-family fee schemes will exclude the poorest of the poor, unless there are some exemptions or subsidy mechanisms. From a national perspective, forcing relatively poor local people to shoulder more of the burden of paying for their own care and letting the nation's elite off the hook, does raise serious equity issues.

Horizontal equity here is paradoxical. In a sense, those with the most complaints will be people living in areas where such schemes are *not* created, and who, therefore, have less service. But this is always an issue with local programs – areas with more local administrative and organizational capacity and community leadership do better. Still, in countries with significant variation in local culture and capacity (e.g., across ethnic or cultural groups), this can be an issue.

5.5.2. Implementability

As with any decentralization scheme, implementability will depend critically on capacity and leadership at the local level. This implies some need for the center to invest in capacity building at the periphery in order to help ensure success. On the other hand, where the theoretical argument does work out in practice, and local people do gain confidence in the arrangement, collections can become easier.

Central governments also need to consider the realistic limits on the organizational and technical capacity available locally. They then need to supply the missing expertise, either temporarily or permanently, depending on what expectations are realistic in any particular situation.

Many villages may not be able to adopt community financing.Villages can be full of internal struggles and divisions by clan, family, ethnicity, religion, and economic status. Any joint community effort may be impossible.

5.5.3. Summary

The main advantages of community financing lie in the credibility those initiatives can engender, and the modest, but valuable, level of risk protection they can offer. Some of the biggest gains may not actually occur in the financing realm at all, but

in the increased efficiency and responsiveness of services when providers are subject to effective local control. The relatively good record of tax-financed community-controlled health services for Aboriginal communities in parts of Australia shows the importance of this aspect of the scheme.

Once again, we would also argue that the devil is in the details, and that implementation matters. There has to be attention to building the needed institutional prerequisites – both technical and political. Otherwise, such initiatives amount to shifting a problem that the central government has not been able to solve onto the shoulders of less well-off and less well-equipped regional and local authorities. Attention also has to be paid to those communities that get left behind when widespread "demonstration project" initiatives are undertaken. Far too often, the successful sites are areas that are better off – politically and economically – while places with the least capacity and most need (e.g., tribal areas in India, tea estates in Malaysia) get the least attention.

5.6. Comments

Before we move on, note *what is not part of our categories*. For example, the general revenue ideal type, in our scheme, does not require that there be direct public provision of services – although that is often the case. Instead, within this category, we include are offered by independent providers (Canada), a mix of private doctors and public hospitals (Sweden), as well as systems that are largely publicly operated services (U.K.). Similarly, the private insurance category includes sales by both for-profit and non-profit entities (like American Blue Cross plans) and insurers who pay providers in all possible combinations of methods. Indeed, we presuppose that insurers can even employ doctors and operate hospitals to provide care directly. For example, the Kaiser Health Plans, which sell health insurance in the U.S., are both a non-profit and a direct care provider. We have constructed our categories this way because we want to leave issues of payment and organization to be settled under the heading of other control knobs. We also want to call attention to the range of choice on non-financing matters that are still available even after decisions on financing have been reached.

6. Resource Allocation and Rationing

Once funds are raised, how they will be used and for whom matters greatly in determining who has access to health care, what types of health care are available and accessible, and their quality and quantity. These in turn influence health outcomes and financial risk protection, as well as the allocative efficiency of a nation's health system and total health care costs.

The need to allocate resource arises from a dismal fact: Human wants far exceeds the resources available. Government, private firms and individual households all have to make painful choices allocating scarce resources; we have to choose how much to spend for health care instead of for other meritorious goods, and decisions must be made on how to spend our money to achieve the best possible results. Government makes this choice at two levels. First, there is the tradeoff between spending in one sector and

spending in another. Second, within a given sector, there is the tradeoff between one program and others. For example, one thousand shillings more for primary care means one thousand shillings less for inpatient hospital services (Bitran Asociados, 2000).

However, allocative decisions alone, even when made "correctly," are unlikely to produce the desired outcomes. Money itself does not produce health care. Funds have to be transformed to health care through organized activities. How organizations and individuals produce health care are influenced by the system's macro-organization and incentive structure. These two control knobs, which are explained in the chapters on payment and macro-organization, greatly affect the effectiveness and efficiency in transforming money into health care. Too often technical experts have taken the overly simplistic view that if resources are allocated appropriately, then health outcomes will be improved. For example, the 1993 World Development Report gave such a simple solution for improving the health of the people.

Resource-allocation decisions are complex and difficult, so we next provide a more detailed analysis of the decision-making process, the principles and mechanisms used to apportion the resources, and the difficult tradeoffs nations face.

6.1. Definitions

The apportionment of health resource occurs at two levels – we decide how to distribute resources and for whom at both the population/community and individual level – but the economic resource allocation literature often makes no distinction between the two. This is a reasonable treatment when market price is used as the allocative mechanism; However, under national health service systems, funds for health care are often allocated first by formula to communities, then other means are used to ration health care to individuals. Health insurance also uses a two-level approach: through the design of the benefit package, the fund first allocates to type of service and then other means are used to ration services to individuals. In this book, we distinguish how resources are apportioned to a community/population group or for selected types of services (such as primary care and hospital services) and called this *resource allocation*; we term how scarce health care is distributed to individuals *rationing*. (footnote the distinction made by Roger Evans between allocation and rationing? Ubel, p19 &20)

6.2. Dilemmas in Resource Allocation and Rationing

Even if everyone agrees that health care has to be allocated and rationed, there is little agreement as to what criteria should be used to establish the rules for how funds should be allocated and how services should be rationed. Nonetheless, some nations seem to be more clear and explicit in setting the rules while others are not: Europeans often debate health care rationing (Klein, BMJ and Len Doyal, BMA), people in the United States almost never do.

Many principles and criteria have been used to allocate funds to the community level or to service types; these ultimately depend on the health system's objectives. Most

of these principles try to take into consideration the health needs of a community. For example, the U.K. and Canada allocate funds to communities based on their income and health status. Under insurance financing, the World Bank argues for the selection of health services to be included in insurance benefit packages based on their relative cost-effectiveness.

On the other hand, conflicting values make it difficult for the public to reach a consensus on a set of clear principles that ration health services to individuals (Maynard, A., BMJ, 1996), perhaps because clear rationing rules could literally mean life and death decision for *identifiable* individual human beings. Another difficulty in establishing rationing rules pertains to whose values matters the most – the individual or a group of experts? This is the major difference between the followers of subjective and objective utilitarianism.

6.3. Mechanisms for Allocation and Rationing

Many mechanisms have been used to allocate funds to the community level or to service types, including formulas incorporating health needs of community, relative cost-effectiveness of health services, and price of insurance and health services. Mechanisms used to ration health care to individuals included price, waiting and travel time, no choice of physicians, and lack of basic amenities or professional courtesy at health facilities. Different rationing methods impact on different patient groups. For example, price rationing impairs the poor households' access to health care more than the rich, while waiting time at a clinic impacts high-waged more than a low-waged workers because the "opportunity cost" of the former is greater than the latter.

Once funds are allocated to population groups or service types, the health care has to be *rationed* to individuals because no nation has been able to devote the resources to provide all the health care that will meet all the people's needs and wants. *Price* is a common measure used for rationing, but it has serious equity implications when used for health care. As an alternative, many nations choose to make selected health services free (or nearly free) and rely on waiting time as the device to ration the limited supply. The U.K. is a good example of the use of cost-effectiveness analysis to decide which health services should be limited, with the want thus exceeding the supply. Quantitative measures of preventive and primary care services are produced to meet patients' wants at zero price, but certain expensive surgical and laboratory services are not; these are rationed explicitly by assigning the non-urgent patients to wait an extended period of time for these services. Other means frequently used to ration limited health services to individual patients include no choice of physicians, poor amenities at the health facilities, and extended travel time.

6.4. Principles of Allocation and Their Implementibility

Regardless of which ethical value is being followed, there is wide agreement that the government should give top priority to funding public and merit goods such as public health and prevention programs, including immunization and maternal and child health.

A major difficult arises, however, in apportioning public resources between primary care and hospital services.

Setting spending priorities between primary care and hospital curative services involve tradeoffs between improving health versus providing risk protection. Allocating resources to pay for primary care is more cost-effective in improving health status but does not provide financial risk protection. Meanwhile, allocating resources to pay for inpatient hospital services offer financial risk protection but is less cost-effective in improving health status. This is the major dilemma for low- and middle-income countries which lack public funds to pay for both.

The World Bank and WHO have urged governments to use the cost effectiveness criteria to allocate public funds and design social insurance benefit packages, though effectiveness only takes health outcomes into account. Yet the public clearly wants and demands financial risk protection from expensive medical services. Consequently, most countries find it impossible to follow the recommendations from these international organizations.

In fact, when countries lack the public resources to fund both primary care and hospital services, most find that they have to given greater priority to paying for hospital services. Table 7.2 shows the large share of public health budgets spent for hospital services for selected countries.

Table 7.2. Share of Government Budget Allocated To Hospitals For Selected Countries

Country	Hospital Share
Bangladesh	**61%**
Burundi	**66%**
China	**61%**
Cote D'Ivoire	**46%**
Ethiopia	**49%**
Jamaica	**72%**
Mexico	**58%**
Philippines	**71%**
Somalia	**70%**
Turkey	**63%**
Zimbabwe	**54%**
OECD mean	**54%**

Source: Barnum and Kutzin

6.5. Political Economy of Resource Allocation

Besides the tradeoffs between allocative efficiency to yield the greatest health gains and financial risk protection, governments that are well-intentioned and which function well also wish to allocate their scarce public resources to achieve the greatest

possible equity gains. Such governments will direct their funds to subsidize those who are poor and those who have greatest health needs but lack the ability to pay. To ??? a government can target its resources to intended beneficiaries through various means. It can target its subsidies by income group, by health and socioeconomic status of a community, by class of hospital ward, and by types of services. Alternatively, a government can directly provide free (or nearly free) services to poor communities and the most vulnerable population, such as the disabled and elderly. Several technical approaches have been used to allocate public funds. Some advanced economies such as the U.K. and Canada allocate funds to regions and communities, based partially on the health "needs" of a community. Capital investments are prioritized based on the cost-effectiveness of a technology or service in producing health outcomes.

Formulating financing policy, like other policy decisions, is a profoundly political process. There are several strong stakeholders in the health sector – organized medicine, labor unions, insurance and pharmaceutical industries. Each possesses political resources and roles in the political structure that determine their relative power in shaping the financing policy.

Political scientists such as Alvarex (1991), Marmor (1983), and Reich (1994) have long argued that politics plays the critical role in deciding who has to pay and who receives the benefits. A country's economic and political elite wants to ensure that there are at least a few "world class" institutions in the country in which they can get care. These services tend to be provided by the tertiary hospitals or medical centers. Priority is often given to tertiary hospitals utilizing costly equipment (frequently imported) and serving the economic and political elite of the country. It is common for those national and regional centers (which are often also teaching hospitals) to absorb a very large share of the nation's overall health budget. Furthermore, the most prestigious institutions often have substantial political connections and influence which allows them to defend their interests effectively.

Data tend to show that many nations are not allocating their health resources equitably, and the allocation of tax funds among programs often reduces rather than enhances the equity of health-care delivery. While general revenues are often used to fund free public health care, which is intended to assure equal access for the poor and low-income households, the reality is quite different, and the benefits frequently do not go to those whom the public funds are intended to help. Similar to what Bates (1981) found in his study of agricultural policy in Africa, general revenue financed health services tend to be "captured" by the urban upper and middle class rather than going to the rural and urban poor. Incidence analyses indicate that public expenditures tend to benefit the rich disproportionately in Ghana, Indonesia, Vietnam and Brazil (see Table 7.3).

Table 7.3. The incidence of public health spending in selected countries

		Share of subsidy (%)	
		Poorest quintile	Richest quintile
Sri Lanka	1979	30	9
Jamaica	1989	30	9
Malaysia	1989	29	11
Brazil	1985	17	42
Egypt	1995	16	24
Kenya	1993	14	24
Vietnam	1992	12	29
Indonesia	1989	12	29
Ghana	1992	11	34

Source: Alailima and Mohideen, 1983, Demery et al 1995, Grosh 1994.

This unequal distribution of public health resources in favor of the rich is alarming. While the political economy of these nations may offer an explanation for these outcomes, it also points out a fundamental difficulty of the health sector. Unlike public health programs, personal health services accrue direct and immediate benefits to identifiable individuals. A rational calculas would tell those who are ill how they would benefit from certain distributions of public resources, and their personal tradeoffs would be very different from the statistical calculations of well-intentioned public officials.

7. Conditional Guidance

Financing policy is complicated, because it has to address both the issues of equity and efficiency. The multiple objectives of health systems require painful trade-offs. While just about everyone agrees that health care is a fundamental necessity for human well-being and that our ability to learn, work, achieve our potential and enjoy life depends on our health condition, there is no agreement as to what financing policy would be optimal to assure that everyone enjoys equitable outcomes. So much of the decision depends on the nation's social values.

How social values influences financing strategy can be illustrated with an example where followers of two different ethical principles would reach very different conclusions. From the egalitarian perspective, every person has certain positive rights. A just and fair society, then, has the responsibility to provide sufficient funds to assure that everyone has equal access to adequate health care to better their health or relieve suffering. The egalitarian philosophy dictates that the government takes primary responsibility for developing the most effective and efficient approach to mobilize the necessary funds for every citizen's health care, rich or poor.

In contrast, objective utilitarians do not believe citizens have positive rights to health care. Objective utilitarians whose ultimate objective is increasing economic

growth would view health care in the context of how health care contributes to economic growth. While the government has the responsibility to organize funding for the portion of health care that could most efficiently improve the nation's economic growth, equal access to health care and equal health status is not its concern.

Our review of the financing strategy and choosing the combination of five health financing modalities lead us to a few painful conclusions. The first is, there is no perfect financing modality. Each option has pluses and minuses, but some are more equitable, or feasible and sustainable than others. The second is, the optimal choice for a country may well depend on exactly where that country is with regard to a set of fact and value issues. How important is vertical equity – and really redistributive financing? How much administrative capacity is there to collective the taxes or social insurance contributions? How cynical and being a free-rider-inclined are citizens when it comes to join community-financing schemes? The third conclusion is that details and implementation matter. The horizontal equity of a social insurance system will depend on how small businessmen are treated. The vertical equity of a system that relies on regional taxes will depend on what kind of interregional equalization mechanism is created.

These complexities suggest why we have so many "mixed" or "impure" schemes in the real world. But that does not mean that what everyone is doing is appropriate. It is possible to do better or worse. It is possible to use international experience and theoretical insight to make plausible predictions about your context. And that is just what is required if this control knob is to be adjusted most effectively.

Both international experience and theory offer several principles and guidelines:

- Poverty itself imposes a basic financial constraint. The poor households simply can't afford to pay for health care; they have to be subsidized by the government. Public spending must be better targeted to the poor which is not the case now for most low and middle-income countries.
- For the two billion poor rural populations, private out-of-pocket spending substantially replaced with prepaid community financing schemes. Such schemes could use the current spending more efficiently and effectively, improve the quality of health care, and pool some of the risks to reduce impoverishment due to large medical expenditures.
- Every low and middle-income nation should conduct a systematic review of how health care is financed, and then prepare a coherent and realistic financing strategy for itself, donors and NGOs. Such a strategy will use a combination of financing methods to mobilize funds from different population groups and will pool risks as widely as possible. General revenue, the most flexible and equitable source of fund, will be used to fund public goods, subsidize the poor, promote equitable access to essential health care, and offer incentives for farmers and informal sector workers to participate in prepaid community financing schemes.

- User fees have generated additional funds for public facilities in some countries, and induced customers to be more careful of their utilization. However, many user fees schemes have not net additional funds for health care because of high administrative costs, and when additional funds were produced, they did not necessarily led to better services desired by the patients. Moreover, many user fee systems did not succeed in assuring the poor to have adequate access to public health services although they were designed to remedy this potential problem. In short, user fees systems require careful design and capable administration before they can produce the desirable outcomes.

Nations that want to mobilize a significant percentage of GDP for health care will have to have a coherent and rational financing strategy to fund health care for all the people. The appropriate combination does depend each country's context – social values, economic condition, industrial structure, and administrative capacity. The capacity of those sources in a particular context will depend on many local features of the situation.

Chapter 8
Payment Systems and Their Incentives

1. Introduction

The second of our five control knobs is the payment system. In the financing chapter (Chapter 7), we discussed the ways in which money is mobilized and used; we will now discuss how money is paid out. Every organized effort to mobilize financial resources has to decide *how to pay providers and who to pay.* The organizations that mobilized the funds (government, social and private insurance plans, and community financing schemes) have to decide first how the services should be organized and delivered (see Chapter 9 on Organization), who are the qualified providers eligible for payment (see Chapter 10 on Regulation), and how and how much to pay.

Payments create financial incentives that have a powerful influence over the actions of organizations and individuals. In this chapter, we will look at the financial incentives affecting both buyers and sellers. The amount patients have to pay affects the demand for services, while how the payment is made and the amount of the payment creates the incentive structure for providers.

The performance of any health system depends on many individual actions. Among the five control knobs, payment and regulation (see Chapter 10) have the most direct and immediate effects on the behavior of individuals. Empirical evidence consistently shows that financial incentives are among the most important influences over organizational and individual behavior. Unlike regulation, which relies on the power of the state to force individuals to comply with the regulatory intent, financial incentives create a material gravitational force that pulls on individuals and induces them to voluntarily change their behavior.

In addition to affecting individual behavior, payments also have a big and direct impact on the health system's cost. The total expenditure for health care is the price per unit of service times the volume of service summed for the whole country (Σ price times quantity). However, we explain later that price also affects quantity and modes of treatment provided through supply induced demand.

Deciding on how to pay depends significantly on the organization of financing and delivery of health services. Financing and delivery could be integrated or separated. For example, when financing organizations (government or insurance plans) directly own and manage the providers (hospitals and clinics), they are integrated. Under this arrangement, certain payment methods become more appropriate and feasible such as salary, line budget or global budget. When financing organizations and providers are independent entities and deal with each other at "arms length," then other payment methods become feasible and appropriate such as fee-for-service or capitation. In the past decade, many countries with integrated financing and delivery systems (e.g. UK and New Zealand) moved to separate their financing from delivery so they could use new payment methods and competition to improve efficiency and quality of health care. Social

insurance plans in Latin American nations have done the same. Chapter 9 discusses the organizational alternatives and their potential impacts.

Once the payment method has been chosen, the level of payment has to be decided and adjusted frequently to reflect inflation and changes in production cost. Setting payment levels become the most direct point of contention between buyers and sellers. Providers want payment levels to be as high as possible because they directly impact their income; payers, on the other hand, wish to pay as little as possible because they have to bear the cost.

Incentive structures create moral hazard[19] on both the demand and supply sides (Arrow 1963; McGuire and Pauly 1993). On the demand side, financial incentives affect the behavior of patients: the monetary price a patient must pay at the point of service influences the quantity and quality of services demanded. Similarly, the price patients have to pay influences their decisions about where to seek care and how many services they want. Even when there are no alternative providers, a patient can choose to make a self-diagnosis and perform self-treatment of minor illnesses. Numerous econometric studies have been conducted to examine the impact of financial incentives on patients, the price effect, by estimating the price elasticity of demand, which is a significant part of "moral hazard." These findings help policymakers in designing benefit packages and user fees.

On the supply side, the payment system produces a more significant and complex set of behavioral responses by physicians and hospitals. The incentive structure induces physicians to change their volume and quality of services, treatment modalities, dual practices (practicing in both public and private clinics), the total number of hours worked, and the number of patients treated per hour . For example, if physicians are paid higher fees on a fee-for-service basis, they work more hours and give more units of services per hour than if they are salaried. The incentive structure also affects hospitals' behavior. Hospitals do alter the length of stay, the admission rate, and the quality of services in response to payment changes.

Physicians can also induce demand for their services. The medical knowledge of the seller, the health professional, is far superior to that of the buyers, the patients, and the possession of this highly specialized and complex knowledge accrues autonomy to health professionals in the making of medical decisions. They can positively or negatively induce demand[20] by advising patients on when to come back for another visit, what drugs to take, what specialists to see, and what laboratory tests or surgical treatments to have. .

[19] "Moral hazard" refers to changes in behavior due to changes in financial incentives. For example, on the demand side, health insurance lowers the price that patients have to pay for health services, it tends to alter patients' behavior in two ways: encourage the insured to be more risk-prone as well as demand greater quantity of services when he is ill. The latter often termed as the price effect. On the supply side, when providers are paid by capitation, for instance, and have to bear the cost of a treatment given to patients, they tend to minimize the treatment given.

[20] "Induced demand" refers to the situation in which providers, due to their superior knowledge of medicine, can alter patients' demand for health care by convincing patients to have more or less volume of services, drugs, tests or a different modality of treatment than what the patients would have decided as best for themselves if they had adequate knowledge about the cost-effectiveness of alternatives. Providers can be motivated to induce demand to increase their income and/or to improve their professional standing.

This chapter is organized into five sections. In the next section, we briefly explain how financial incentives have an impact on intermediate and core outcomes. Section III discusses the four key factors to be considered when designing a payment system; in section IV, we describe and analyze the principal payment methods. Lastly, we present the conditional guidance.

2. Impact of Payment on Outcomes

Payment creates an incentive structure that affects individual behaviors, for both buyers and sellers. The price that patients pay influences the volume of services they use, which in turn, affects patients' health status. For instance, a patient who pays out-of-pocket for expensive services may choose to forgo care because she has to borrow to pay for the services – a decision that may adversely affect her health. Another example is the situation in which a patient has insurance that pays for prescription drugs. When he visits his physician, he demands a particular brand-name drug because he has read an advertisement for it, even though there is a generic drug that is therapeutically equivalent and costs much less. Here, satisfying the patient's demand reduces the efficiency in producing health, and raises health care costs, which are intermediate outcomes. Eventually, the public may become less satisfied with the health system because it's so costly, a core outcome. Figure 8.1. illustrates these relationships.

For providers, the payment method determines the amount of financial reward they receive and the amount of financial risk they bear. In response to these incentives, the provider could improve or reduce efficiency and quality, thus impacting health status and influencing the cost of health care and the patient's satisfaction. For examples, when physicians are paid on salary instead of fee-for-service, they see fewer patients each hour and provide less personal amenities. When a government pays hospitals a fixed amount set in advance for each patient admitted, the hospital would want to admit as many profitable cases as possible, and avoid the money-losing cases. The hospital would also want to operate efficiently so it could earn a surplus – the difference between the cost of treating each patient admitted and the amount per admission received from the government – by discharging patients as soon as possible to reduce cost. However, if patients are discharged too early, their health may be adversely affected.

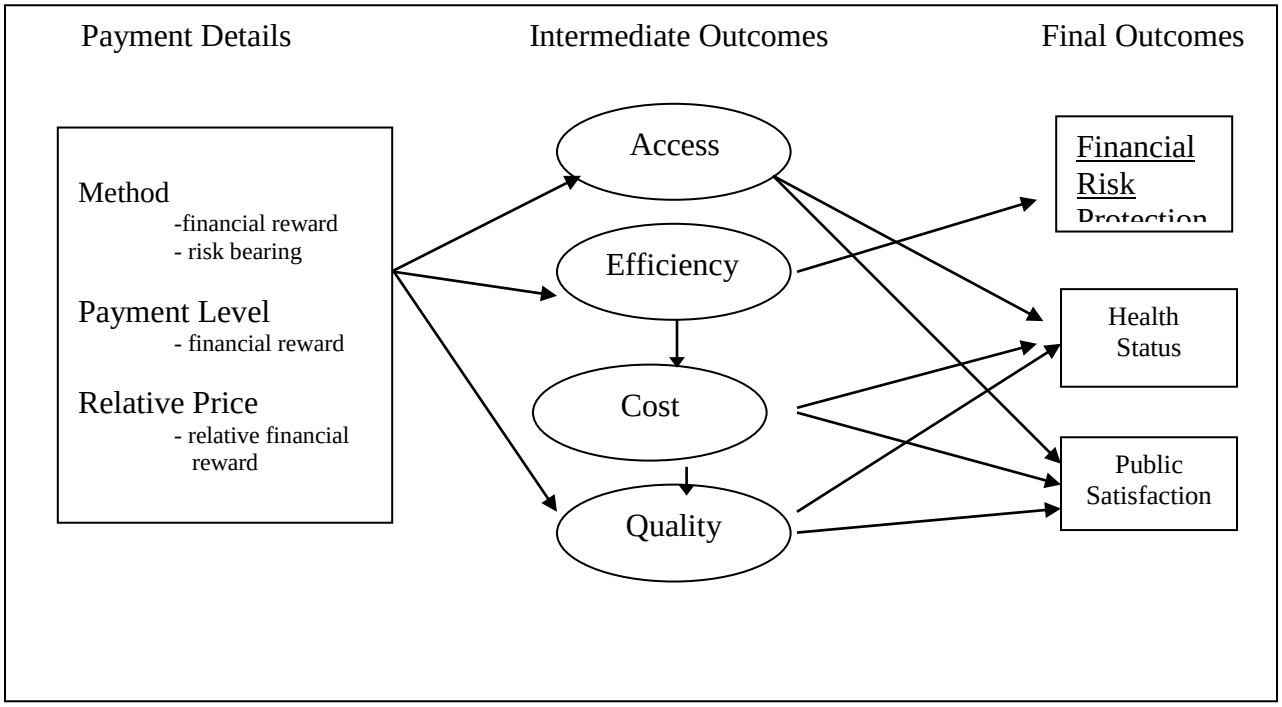
The relative price of different services also impacts providers' behavior. When a patient can be treated with different services that may produce the same outcome, the relative price difference of these services influences which service is provided. For example, when the payment for caesarian sections becomes more lucrative than normal delivery, physicians are more likely to deliver babies by caesarian section. Studies have confirmed this prediction, showing that higher relative payment levels for caesarian sections lead to higher rates of caesarian sections . Such behavioral changes due to relative prices can impact the quality and cost of health care.

Differential payment rates by different payers could have significant effects on patients' access to health care and their quality. For example, many countries have several insurance plans offering health insurance and the plans pay different amount for the same service rendered by the same provider. Then the provider may discriminate against patients insured by plans that pay lower rates. This discriminatory practice can be profound in countries where government-employed physicians can have outside private practices, and they charge high prices so the lower-income households can't afford to pay.

Prices, in a marketplace, give crucial signals to both consumers and providers. Payment level serves similar functions. On the supply side, the prices paid influence how many providers will enter the market and how services will be produced. On the demand side, patients decide which goods to purchase and how many, based on the amount they have to pay. It's the key economic allocation mechanism to ration scarce resources in an economy – they determine the production, distribution and consumption of health services, drugs, and capital funds. In sum, payments and their incentives play a critical role in health system performance.

In payment systems, we distinguish between the method, level and relative payment levels. Each of these influences the behavior of physicians, hospitals and pharmacists through financial reward and risk bearing. Figure 8.1. illustrates the relationship between payment, intermediate outcomes and core outcomes.

Figure 8.1. Relationship between Payment and Outcomes



3. Design Decisions

In designing a payment system, the reformer must keep in mind the fundamental considerations described in section II: its powerful effect on the individual behavior of both buyers and the sellers, it creates two types of incentive – financial rewards and levels of risk bearing, and the importance in getting the relative payment levels right.

In this section, we turn our attention to the four basic decisions that have to be made in designing a payment system: the unit of payment, level of payment, determination of rates retrospectively or prospectively and adjustments of rates over time.

3.1. What payment method to use and how to define the unit of service to be paid?

Health financing institutions have to decide how to pay providers, given a macro-organizational arrangement. The paramount question is what payment method to use and how to define a unit of service. Payment method and unit of service are closely linked. The definition and coding of service units impacts the total supply, and the efficiency and quality of health services. For example, health services can be bundled or unbundled. When services such as tests, drugs and procedures for a given disease are bundled together into one payment (such as per admission or per day) hospitals are encouraged to be more efficient. At the same time, there is the risk that some services, such as tests, might be under-provided.

There are several choices in payment method and defining the unit of service. We summarize the options in Table 8.1. The incentives created by each payment method and their impacts are presented in Section IV – Payment Methods.

Table 8.1. Unit of Payment and Payment Methods Commonly Used

Payment Method	Unit of Payment	Physicians & Other Health Professionals	Hospitals and Other Provider Institutions
Fee-for-service	Each Item of service	**X**	**X**
Salary	Units of time worked	**X**	
Capitation	Each person registered	**X**	**X**
Per diem	Each day		**X**
Per admission	Each admission		**X**
Case-mix adjusted admission	Each admission by treatment category		**X**
Line item budget	Each budget line		**X**
Global budget	Each hospital		**X**

3.2. How to set the level of payment?

Once the unit of payment has been decided, the level or the rate for each unit of service has to be determined. The rates do have significant impacts. If the payment rate is set too low, it affects the quality and supply of services. If it is set too high where providers can earn a handsome profit, the resources will be used inefficiently and total health expenditure will be higher. Equally important, the higher earnings could induce excess supply.

Most countries try to set the rates based on some appropriate empirical evidence. For example, under fee-for-service, the rate can be based on the providers' charges; however, this assumes the market is reasonably competitive and that charge reflects the cost for producing that particular service efficiently. Because of serious market failures, seldom health service market produce charges that reflect the minimum production costs. Most large social insurance realized that they should not base payment rates on the amounts providers charge. Private insurance plans, on the other hand, do usually pay providers according to their charges, because they lack the capability to set prices otherwise.

Since charges are not likely to be the appropriate basis to set rates, a country or an insurance plan may decide to set the payment rate based on the actual resource costs. However, it faces great difficulties in determining what constitutes a "reasonable" resource cost for every service. Each provider has its own production cost, some more efficient than others. Furthermore, clinics and hospitals in the same region calculate their costs in different ways.

Collecting cost data is not a simple matter. The country must already have a uniform cost accounting system in place so the costs reflect the same set of activities. For example, the DRG payment method needs *uniform* data on the costs of diagnosing and treating each major disease.This requires hospitals to have uniform accounting and cost accounting systems, along with national uniform clinical record systems that standardize the coding of diseases and treatments. These systems are complicated and expensive to establish. Most low- and middle-income countries do not have these systems in place.

For hospitals managed by governments or insurance plans, most of them have been paid based on line item budget or global budgets, set in advance. In these situations, often the historical budgets became the basis in setting the future budgets.

To overcome these difficulties, several other approaches have been used. Many organized payers negotiate the rates with providers. Some payers use the "most favored nation" approach to negotiate the rates, which involves selecting the charges made by known low-cost providers or charity providers and using them as benchmarks. Germany has developed a negotiation process that was successfully in setting reasonable payment rates. It relies on a negotiation process where the representatives of those who have to pay the tax/premium/fees, negotiate directly with those who will receive the payment, the health care providers. The government tries not to interfere in these negotiations.

In many situations, payers can use competitive bidding to determine the payment rates. But while competitive bidding can at least theoretically produce price levels that are close to the bidders' actual costs, providers often object strongly to this approach because the less efficient among them would lose. Furthermore, competitive bidding cannot produce satisfactory results unless the process is well designed and the appropriate information is given to all potential bidders. And when the government manages a competitive bidding process, politics often intrude. Consequently, payers can end up paying high rather than reasonable prices.

As a result of these difficulties, many nations have moved away from the fee-for-service payment method because it's difficult to set rates for the enormous number of service items. For example, the insurance programs in the United States officially recognize more than nine thousand service items for payment, China recognizes close to four thousand items. Instead, whenever possible, governments and health insurance programs are relying more on capitation methods to pay for physician services, and global budgets to pay for hospital services.

Setting payment levels for physicians have been extremely contentious because few objective criteria can be used to set the reasonable income for medical professionals. Their potential earnings in private practice is not a good reference point because physicians possess monopolistic power to set high prices in the marketplace. Sometimes, nations use the income of other highly trained professionals as benchmarks. Economics can provide an objective indicator for the appropriate payment level. We can judge the reasonableness of a payment level by whether the payment rate has been able to balance the supply and demand for physicians. However, physicians in most countries have not been willing to accept this criterion.

3.3. Will the payment level be set prospectively or retrospectively?

The payment level (i.e. rate) is either determined prospectively at the beginning of a year, or retrospectively at the end of the year when the actual costs for delivering services are known. The approach chosen affects the provider's concern for efficiency and controlling the cost of production. Under prospective rate setting, the payment level is set in advance for each unit of payment, regardless of the actual cost incurred when the patient is treated. The providers therefore have the incentive to operate as efficiently as possible to earn a surplus. Under retrospective rate setting, the providers have the incentive to spend as much as they can to raise their cost base and get reimbursed at the end of year.

For example, when the per admission payment is set prospectively, hospitals have an incentive to minimize the cost for each admission when the payment level is set prospectively. On the other hand, under retrospective payment, the amount per admission depends on how much the providers have actually spent. Prospective rate setting encourages greater operational efficiency than retrospective rate setting. However, retrospective rate setting reduces the likelihood that providers would under-treat patients.

3.4. How is the payment level updated?

To keep the payment level current requires annual updating. Some methods, such as charge-based and retrospectively cost-based payments, are self-updating as either the charge or cost increases. However, updating the payment level under prospective rate setting poses a challenge in attempting to reflect future price levels, and tension grows between the payer and the provider. Payers and individual patients who pay out-of-pocket, want the payment level to be kept as low as possible, while providers want the level to be as high as possible. To minimize this tension, the level is sometimes indexed to general price or wage levels. Some countries index the rate of increase to less than the general price increase, to encourage providers to improve operational efficiency. Hospital budget payment, based on prior year's actual expenditures, creates the same perverse incentive as setting rates retrospectively.

All the above decisions have to be made under political constraint. Even when a country relies on private insurance financing, politics still plays a major role. Stakeholders often ask government to intervene in their payment negotiations. These interventions can be direct, by regulating the payment level, or indirect, by penalizing private insurance plans for withdrawing from a particular market when they refuse to pay the rate demanded by providers.

Establishing a payment system is an extremely contentious affair: It determines the income level of hospitals, physicians, nurses and pharmacists and how much patients have to pay at the point of service, and correspondingly determines the amount of taxes or insurance premiums people have to pay. When the government sets the payment method and level, all stakeholders will engage in an intensive political struggle: payers want to minimize their financial burden and providers want to maximize their income. In the end, the decision must be politically acceptable.

While payment rates can be set by a politically autonomous agency to insulate the decisions as much as possible from political influence, politics still play an important role.

4. Payment Methods

In this section, we examine the incentive structures that operate in two important stakeholders of health care: health care providers and patients. In section A, we first present the payment methods commonly used around the world and summarize how each method affects the behavior of provider organizations and health professionals. We next examine how incentive structures affect the technical efficiency and quality of health services. Finally, we summarize recent empirical evidence to assess the strength of the incentive structure. In section B, we do the same for the financial incentives that operate on patients.

4.1. Health Care Providers

A payment system for health care providers has three parts. The first concerns the method of payment and determines who is bearing risk. The second part concerns the amount paid per unit and defines the magnitude of financial compensation. The third part, which is the relative amount paid for different services, sets the different opportunity costs to the provider within his or her practice.

The method of payment determines the unit of service for which a payment will be made. It ranges from all the services rendered by a provider over a given period of time (e.g., a month of salary), to every single act performed by a professional, such as an injection or a consultation (e.g. fee-for-service). Physicians can be paid on the basis of fee-for-service, capitation, or salary. The method of payment for hospitals include fee-for-service, per diem, per admission, case-mix adjusted per admission (DRG), or global budget. In practice, societies rarely use only one single payment method; instead, two or more forms are combined and tailored to the particular situation and requirements of a specific health care delivery and financing system.

Each of these payment methods creates different financial rewards for providers and shifts financial risk to different players in the system. Because providers decide on the method of treatment and the quantity of services and drugs, the incentive structure affects cost, efficiency, and quality of health services.

By shifting the financial risk for treatment costs between payer and provider, payment method influences the behavior of the provider in many ways. For example, a payment method such as capitation, which shifts the financial risk to the provider, motivates the provider to minimize services.

Below we provide a brief explanation of the major payment methods for health providers, and summarize the findings of some of the most recent empirical studies on the impact of these payment methods. For clarity purposes, we divide the payment methods into two payee categories (physicians and hospitals) although fee-for-service and capitation apply to both.

4.1.1.Payment Methods for Physicians and Other Health Professionals

Physicians who work for health care organizations normally receive financial compensation in three ways: salary, salary plus bonus or capitation. In some situations, physicians can receive compensation through a combination of these three methods (e.g. base salary plus capitation). In China, for example, physicians can earn a monthly salary and a "bonus" based on the number of patients seen. In the US, physicians employed by clinics earn a salary and a bonus based on performance, measured by resource utilization and patient satisfaction. These assessment criteria can be either individual or group-based.

Over the last decade, a growing interest has emerged on the role of financial incentives in health care organizations. Inspired by other industries, health care managers have designed financial incentives in order to induce more cost-reducing behavior on the

part of physicians (these financial incentives often run counter to other professional and legal constraints however). Recent research on health services and organizational theory has highlighted the importance of three factors in the design of organizational incentives: proximity, intensity, and interaction (Harshberger 1999).

"Proximity" refers to how much risk the physician bears in his/her relationship with the patient; capitation, for example, represents a high proximity incentive structure. Physicians' incomes are directly linked to their practice of medicine, and "intensity" represents the magnitude of the responsibilities and incentives facing the individual physician. The questions that determine the intensity level of the incentives include what services are included in the capitation rate? What is the maximum potential loss or gain for the physician? When are bonuses distributed? (Pearson et al 1997) The last factor, interaction, refers to the behavioral dynamic among physicians who are paid differently.

4.1.1.1. Fee-for-Service (FFS)

The unit of payment consists of individual visits for clinical activities such as injections, laboratory tests, and x-rays. This payment method gives providers an incentive to perform more services. This is the only form of payment under which the provider does not have any incentive to select healthier patients; in fact, the opposite is true. Under FFS, the provider bears no risk for the cost of treatment; the payer, the insurer or the patient is entirely at risk for the cost of care. Theoretically, patients and third-party payers have a reason to question the cost of services and the need for additional services, and to negotiate lower payments; but in reality, however, patients and third-party payers can seldom negotiate effectively because physicians and hospitals have much greater professional power.

Studies in African countries have found that shifting from budget payments to FFS increased the volume of services and their cost by fifty percent . Comparing resource utilization under two provider payment methods (FFS and capitation) in Thailand, Yip et al (2001) found a significant difference in the average length of stay, drug charges per admission, and lab costs per case when comparing the two methods. Under FFS, resource utilization by the providers was consistently greater. A study in the Unites States (Krawelski et al, 2000) found similar results. Costs were significantly less under medical group practices paid by capitation than those paid by FFS.

Although both theory and evidence point to the inflationary nature of the FFS payment method, it is the still most widely used method in developing countries to pay private sector hospitals, health centers, and individual practitioners. This could be explained by the fact that most governments of developing countries have taken a "benign neglect" position toward private sector providers. Thus, providers establish their own unit of payment and level and choose FFS since it is the easiest, most profitable and most flexible method.

4.1.1.2. Capitation

The unit of payment is defined on a per-person basis. A payment is fixed to pay for all the services that a person may use during a period of time, such as a month. The most common capitation payment is when a fixed rate is paid to a general practitioner for each patient registered with him for that month, regardless of the services required by or rendered to the patient. This payment can vary by the patient's age, sex, and health status. Capitation has been used to pay for primary care (capitated to general practitioners), for all specialist services (capitated to multi-specialty clinics) and for inpatient services (capitated to hospitals). It is also possible to bundle the primary, specialty and inpatient services into one capitation rate that is paid to one single entity (e.g. GP fundholders[21]) that takes responsibility for all services.

Capitation payments transfer most of the financial risks to the providers; these vary, according to the specific contract terms and organizational arrangements. For example, many Colombian insurance plans contract general practitioners for primary care services and pay them by capitation . On the other hand, the British National Health Service (NHS) contracts GP fundholder groups for all primary care, specialist outpatient and hospital inpatient services, paying them by capitation .The fundholder groups may in turn, contract hospitals for both specialty services and inpatient services, and pay them by capitation or possibly, FFS. In comparison, social insurance plans in Thailand contract hospitals for all primary care, specialist and inpatient services, and pay them a capitation rate. The hospitals then contract with GPs for primary care and pay them a capitation rate . The incentives under the capitation method vary according to the specific services being covered. In the NHS example, GP fundholder groups bear the financial risk for all care; general practitioners therefore have an incentive not only to limit primary care and drugs but also diagnostic, specialist and inpatient services, in order to be financially solvent or to earn a surplus at the end of the year.

Numerous studies have examined how incentives under capitation have had an impact on provider behavior. Iversen (2000) evaluated the impact of capitation on Norwegian GPs' referral decisions. He found a 42% increase in the rate of referral from general practitioners to specialists after Norway introduced a new, experimental payment system in which GPs are contracted only for primary care and are paid by capitation. Under this arrangement, Norwegian GPs referred patients more frequently to specialists, reduced the average number of visits per registered patient, and increased the number of registered patients in their panel. In another recent study, Bitran (2001) found that in response to rising health care costs, Argentinean social insurance plans have been moving away from direct provision and towards the purchasing of health services via capitation. After shifting to capitation in 1997, one insurance plan experienced a drop in the number of hospitalizations per 100 beneficiaries from 2.83 in 1997 to 2.6 in 1998.

Under capitation, providers may choose to accept only healthy and less complicated patients, otherwise known as "risk selection." Newhouse (1996) examined

[21] The term "GP fundholder" originated in the United Kingdom and refers to those general practitioners (GPs) who receive a fixed amount for a set of comprehensive services that goes beyond what the GP would provide directly, such as specialist services and laboratory tests.

the trade-off between risk selection and efficiency of production under managed care plans' capitation payment in the United States. He concluded that capitation payments encourage providers to become more efficient in their use of resources but they may, on the other hand, it create risk-selection problems: When providers received a fixed payment for every registered patient, they tended to select the healthier ones to minimize their exposure to risk.

4.1.1.3. Salary

The unit of payment is based on a time period that employed physicians are at work, regardless of the number of patients seen, volume of services, or cost of services provided. Physicians paid by salary bear little financial risk and have no incentive to minimize or maximize the number of visits. They may, however, be concerned about the amount of time they spend at work. In developing countries, their employer, usually the government, bears the financial risk, and has incentives to minimize costs and maximize efficiency; the employer may therefore, ask each physician to see many patients per hour. Furthermore, there is often tension about what the salary should be. In developed countries, the employers, such as prepaid medical plans, often use incentives to induce physicians to be cost-conscious. For example, many prepaid health plans in the US offer bonuses to physicians if the health plan operates under budget.

A recent study reviewed 23 published papers on salary payments and their effects on providers' behavior in developed nations. Gosden et al (2001) found some tentative evidence that salary payment is associated with lower productivity when compared with fee-for-service and capitation payment methods. Specifically, they found an association between salary payment and (l) a reduced number of services per patient; (2) a reduced volume of patients per physician; (3) longer patient visits; and (4) greater degrees of preventive care compared to fee-for-service. Other anecdotal evidence indicates that salary payment does not encourage physicians to be "responsive": the lack of financial incentive dampens their enthusiasm towards satisfying the concerns of patients and fellow physicians (Harshbarger 1999).

Each of the different payment methods creates a particular set of risks for providers. In Table 8.2. we summarize the payment methods, and then show how each method produces a financial reward or risk for the physician and the payer, and the likely physician response to the different payment methods. The payer could be a patient or an insurance plan.

Table 8.2. A Summary of Payment Methods for Physicians: Their Financial Risk and Incentives

		----------Risk borne by:----------		----------Provider Incentives to:----------			
Payment Method	basket of services paid for	payer	provider	increase number of patients treated or registered	decrease number of services per chargeable units of care or consultation	increase reported illness severity	select healthier patients
FFS	Each item of service and consultation	all risk borne by payer	no risk borne by provider	yes	no	yes	no
Salary	One week or one month work	all risks	no risk borne by physician	no	N/A	N/A	no
Salary and Bonus	bonus based on # of patients	salary portion	bonus portion	yes	N/A	N/A	yes
Capitation	all covered services for one person in a given period	amount above "stop-loss" ceiling	all risk borne by provider up to a given ceiling (stop-loss)	yes	N/A	no	yes

Sources: Prepared by William C. Hsiao 1997, modifying data from WHO1993, Bodenheimer and Grumbach *JAMA* 1994

4.1.2. Payment Methods for Hospitals and Provider Institutions

4.1.2.1. Per admission

The unit of service is defined as an admission and a fixed amount covers all the services given in the episode of hospital stay, regardless of the actual volume or cost of services provided. This method transfers a portion of the financial risk to the provider. Generally, as the number of services bundled together increases, the financial risk borne by providers increases as well. When a hospital is paid a fixed amount for each admission, which includes all services connected to that admission regardless of the actual amount of services provided or their costs, the incentive for the hospital is to reduce the lengths of stay and the number of services given. The case payment method also gives hospitals incentives to select patients with less severe illnesses and to admit as many patients as possible.

4.1.2.2. Case-mix adjusted per admission

To remedy some of the shortcomings in per admission payment such as hospitals may select to admit only patients with less severe illnesses that cost less to treat, this

case-mix adjusted payment separate admissions into broad disease and treatment categories based on the costs required to treat the patient. Diagnostic related grouping (DRG) has become the most widely used method to adjust for the case-mix.

A recent study evaluating the impact in Italy of one type of case payment, the DRG-based payment, Louis et al (1999) reported a 21.1% decrease in hospital bed days after the introduction of this payment system. Another study examining the impact of the DRG payment method on hospitals in the United States (Gilman 2000) found evidence of a "selection" effect: hospitals reacted to low DRG-based payment rates by selecting healthier patients to treat. These recent findings are consistent with the findings of an earlier synthesis study (Ellis and McGuire,1993).

4.1.2.3. *Perdiem*

The unit of payment is defined on a per-day basis. A fixed rate payment per day of outpatient care or hospitalization is paid to the provider. This method bundles services by hospital day or outpatient care day, regardless of the actual services given and their costs. This method is commonly used for payment to hospitals. This fixed rate per day gives hospitals incentives to be more efficient and to reduce tests and procedures. The hospital, which now bears a larger portion of the financial risk, has an incentive to limit patient services for each day and to keep patients in the hospital longer. As a result, per-diem payments encourage hospitals to have a high occupancy rate and expand their bed capacity. Prudent payers find it necessary to perform utilization reviews to control the unnecessary lengthening of hospital stays.

Studies have found strong statistical correlation between per-diem payment and average length of stay. For example, Rodwin and Okamoto (2000) examined the impact of per-diem payments introduced by the Japanese authorities and found that the average hospital length of stay (ALOS) in Japan are three times longer than those in the United States for the same diagnostic categories. In the case of appendicitis, the ALOS in Japan was 9.8 days while in the US, where hospitals are often paid on a per-case basis, the stay was 3 days. Financial incentives, among other reasons, induced hospital managers to hold onto admitted patients as long as possible. Similarly, in the new Republic of Slovakia, Langenbrunner et al (1999) reported ALOS of 7.5 days for childbirth under a nationwide per diem payment system.

4.1.2.4. Line-Item Budget

The unit of payment is by expense category (e.g. salary, supplies, transportation, drugs) for an organization. The amount budgeted is usually based on the class of that facility (e.g. class 1 or class 2 facility), staffing positions approved, and past budget amounts given. Once the funding agency (e.g. Ministry of Finance, private prepaid health plan) has approved the budget, the provider (e.g hospital, clinic) has little discretionary use of funds across different budget items. Ministries of Health, for example, determine the budgetary allowances of each individual item in a hospital budget (e.g. salaries, maintenance) and require the hospital administrator to stay within the limits

of the line item allocations. At the end of each year, budgetary allocations are discussed and negotiated between the central authorities and the hospital directors. This type of budgeting provides an incentive to the hospital directors to overestimate budgetary needs. It does not financially reward efficiency in the delivery of health care.

Langenbrunner (1999) studied the impacts of line item budget payment systems in the former Soviet Union and former socialist countries in Eastern Europe. He made several general conclusions about line-item budgeting as it was practiced: (1) little incentive existed for facility directors to be cost-conscious and innovative; (2) facilities tended towards underproviding health services; (3) little attention was paid to health outcomes and patient satisfaction; and (4) no real incentive existed to downsize the level of fixed resources (i.e. staff and facilities). Given the shift away from centrally managed health delivery systems, these countries have been experimenting with new payment systems such as per diem, case payment, capitation and global budgets.

4.1.2.5. Global Budget

This is an all-inclusive operating budget set in advance, which provides a total sum and allows flexible use of the funds inside it. This method is used for payment to organizations such as hospitals and health centers, and the organization must meet certain production targets such as number of bed days or number of visits. If the targets are not met, the organization is penalized. A global budget represents "the highest form of bundling services: every service performed on every patient during one year is aggregated into one payment."

Under a global budget payment system, directors of hospitals and health centers have the incentive to control their expenses while attaining their production targets; however, they can also respond perversely. For example, if the number of patient days is a production target, the hospital can prolong patient stays unnecessarily. German hospitals operated on global budget payment system and used the number of "expected bed days" as the production target, Which is part of the reason why Germany has the longest average length of stay among major developed nations. Its average length of stay is three times of the US, where hospitals are mostly paid on a per DRG case basis.

Global budgets have one major operational difficulty. How much should a hospital be paid when it exceeded its production target and how much should it be penalized when it fails to reach the target. This issue becomes complicated, because hospitals have both fixed and variable costs when treating patients. In the short run, the average cost for treating a patient is greater than the marginal cost because of the fixed costs. If global budgets are used with a production target, it is similar to paying hospitals on a per-diem basis once the target has been met. The per-diem rate is the average cost per day for that hospital. When a hospital exceeds its production target, and if the hospital receives payment for the additional days based on average cost, the hospital can increase profits until it is fully occupied. Paying the hospital its additional marginal cost when it exceeds its production target, and reducing payment by the marginal cost when it fails to meet the target, can solve this problem.

Germany is one country that implemented this approach. During 1993-1995, it used the number of "expected bed days" as a target volume (Hoffmeyer et al 1994). When hospitals exceeded this pre-set target, Germany's "sickness funds" would pay hospitals only up to 25% of the average cost per day. When the target was not met, hospitals would be penalized by 25%. So, even if a hospital does not meet its target, it can nevertheless cover the fixed portion of its overall cost and stay in operation.

In Table 8.3. we summarize the payment methods for hospitals, the incentive structures created for the payers and the hospitals, and the likely response of hospitals to the various payment methods.

Table 8.3. A Summary of Payment Methods for Hospitals: Their Financial Risk and Incentives

		----------Risk borne by:----------		----------Provider Incentives to:----------			
Payment Mechanism	basket of services paid for	payer	provider	increase number of patients	decrease number of services per payment units	increase reported illness severity	select health-ier patients
FFS	each item of service and consultation	all risk borne by payer	no risk borne by provider	yes	no	yes	no
Case-mix adjusted per admission (e.g. DRG)	payment rates vary by case	risk of number of cases and case severity classification	risk of cost of treatment for a given case	yes	yes	yes	yes
Per admission	each admission	risk of number of admission	risk of number of services per admission	yes	yes	no	yes
Per-Diem	each patient day	risk of number of days to stay	risk of cost of services within a given day	yes	yes	no	no
Capitation	all covered services for one person in a given period	amount above "stop-loss" ceiling	all risk borne by provider up to a given ceiling (stop-loss)	yes	yes	N/A	yes
Global Budget	all services provided by a provider institution in a given period	no risk borne by payer	all risk borne by provider	no	N/A	N/A	yes

Sources: Prepared by William C. Hsiao 1997, modifying data from WHO1993, Bodenheimer and Grumbach *JAMA* 1994

4.2. Patients

One way payment systems affect patients is that they define how much patients must pay for a service. Two methods are used to create the incentive, one direct and one indirect. Both, however, ultimately reduce the price patients have to pay when obtaining health care. This reduction in price achieves more equitable access and decreases the chance of patients becoming impoverished by large medical expenses.

Under the direct approach, nations subsidize the cost of health care provided at public facilities. This practice reduces the price faced by patients when they utilize services, resulting in increased demand. Under the indirect approach, government can subsidize the premium consumers have to pay to buy insurance, resulting in a lower price for insurance; more insurance will therefore be purchased, *ceteris paribus* (assuming other factors remain constant). This results in a loss of allocative efficiency, which is particularly valued by subjective utilitarians. In addition, excessive insurance causes greater moral hazard.

For nations offering free public health services, some have imposed user fees to generate revenue and reduce moral hazard. Those nations who finance health care through insurance often incorporate deductibles, co-payments, coinsurance, and payment ceilings. All of these are intended to reduce moral hazard and each creates a different set of incentives and risks for the patient.

"Deductibles" are initial payments made by the patient before the insurance policy begins to cover incurred expenses. The deductible level is set for either individuals or families and is sometimes based on the level of income. If the deductible is set too high, an individual would be deterred from buying the insurance policy since he would bear a large portion of the financial risk.

"Co-payments" are fixed payments made by the patient every time he visits a physician or each day he stays in a hospital. These payments are usually predetermined and clearly stated in the insurance policy. "Co-insurance" refers to patients being reimbursed by insurers for a certain percentage of the service price. In this case, suppose an insurer sets a co-insurance level of 30%–the patient would only be responsible for 30% of the amount charged for service. If the service price increases over time, the insurer will still pay 70 % of the new price. The major advantage of co-payment and co-insurance is the spreading of risk between the insured and the insurer.

"Payment ceilings" are a maximum amount of money an insurer will pay per patient per year. If a patient buys an insurance policy, the insurer may only cover that patient up to a predetermined amount – the payment ceiling. A disadvantage of these arises when a patient suffers from a catastrophic illness and needs very expensive services. If the cost of the care exceeds the payment ceiling, the patient alone bears the financial risk. If the patient does not have adequate resources, she may be denied access to appropriate care.

Table 8.4 summarizes the incentive structures that affect patients and insurance plans under free care, full user fees, and various insurance benefit designs. As shown in the table, free care or a full user-fee option leads to an unequal distribution of risk between the insurer and the patient. Under free care, the insurer bears all the risk and vice versa under a full user fee. The other benefit designs spread the risk and generally tend to reduce the amount of health care demanded by the patient.

Table 8.4. Financial Risk and Incentives on the Patient and Insurer Under Free Care or Insurance Covered Benefits

	----------Risk borne by:----------		Incentive for:
D	Payer (e.g. insurance or government)	Patient	Patient
Free	All	none	increase demand
Full user fee	None	all	reduce demand
Deductible	amount above deductible	amount up-to-total deductible	reduce demand until deductible amount reached, then increase demand
Fixed copayment per visit	full charge minus copayment	copayment	reduce demand for visits
Coinsurance (% of charges)	(1-X)% of charges	X% of charges	reduce demand (depending on percentage of coinsurance)
Ceiling on amount paid by insurance	amount below ceiling	amount above ceiling	reduce demand when ceiling is exceeded

William C. Hsiao 1997

Moral hazard has been a major theoretical concern in the public debate over government subsidies towards the purchase of health care. As the public exhibits greater moral hazard, health services markets become less efficient because of the price effect. Its relative importance can be determined empirically by studying the how changes in the price of insurance coverage affect the demand for health care.

We summarize this price effect or the "elasticity of demand"[22] in Table 8.5. below. On average, elasticity is not high, which implies that the distortion may not be serious. However, the elasticity of low-income households is much higher. As to the comparative values between nations, while the elasticity of demand for the US might seem to be somewhat lower than the developing countries', such a conclusion is unwarranted because their difference could be due to the differences in data collection and statistical

[22] Elasticity of demand refers to the percentage change in demand for a good or service in response to a percentage change in the price of that good or service.

methods used.[23] The results for developing countries consistently show that elasticity is higher for lower income households and for children.These elasticity values imply that requiring patients to pay direct out-of-pocket payments, such as user fees, have greater adverse impacts on the poor and on children.

Table 8.5. Elasticity of Demand for Health Services: Overall, by Income Quartile and by Age

Study (Year Published)	**Location (Year of Data)**	**Results**		
Jimenez (1989)	Ethiopia (1985)	Overall:	-0.05 to -0.50	
Jimenez (1989)	Sudan (1986)	Overall:	-0.37	
Yoder (1989)	Swaziland (1985)	Overall:	-0.32	
Gertler & Van der Gaag (1989)	Cote d'Ivoire (1985)	Rural Hospitals		
		Income Quartile	Adults	Children
		Lowest	-0.47 to -1.34	-0.65 to -2.32
		Second	-0.44 to -1.27	-0.58 to -1.98
		Third	-0.41 to -1.18	-0.49 to -1.60
		Highest	-0.29 to -0.71	-0.12 to -0.48
Gertler & Van der Gaag (1989)	Peru (1985)	Rural Hospitals		
		Overall:	-0.57 to -0.50	-0.41 to -0.81
		Income Quartile	Adults	Children
		Lowest	-0.57 to -1.36	-0.67 to -1.72
		Second	-0.38 to -0.91	-0.48 to -1.20
		Third	-0.16 to -0.37	-0.22 to -0.54
		Highest	-0.01 to -0.04	-0.03 to -0.09
Sauerborn et al (1994)	Burkina Faso (1985)	Overall:	-0.79	
		Age Groups:		
		<1	-3.64	
		1-14	-1.73	
		15+	.0.27	
		Income Quartile		
		Lowest	-1.44	
		Second	-1.21	
		Third	-1.39	
		Highest	-0.12	

Source: Reddy, Sanjay and Vandenmoortele, Jan. *User Financing of Basic Social Services: A review of theoretical arguments and empirical evidence.* Office of Evaluation, Policy and Planning. UNICEF, New York, 1996.

[23]The findings from the RAND study came from a social experiment where complete and accurate time-series data were gathered and analyzed. Meanwhile the other studies used data from one-time household surveys.

5. Conditional Guidance

On the buyer side, price affects a patient's purchasing decisions. The average value of elasticity of demand, however, is not high. The loss of economic efficiency would not be tremendous. However, the elasticity values are much higher for low-income households, which should be taken into account when determining how much patients can pay out-of-pocket for health services and who should pay for them. Requiring patients to pay some user fees for services and drugs when their elasticity of demand is high seems to be a sound practice. For example, when drugs are made free to patients, it often leads to large wastage, including cases of drugs being taken home and never consumed. When designing the mechanism to use by which patients will pay, flat user fees, including co-payments, are the easiest to administer. For insurance plans, deductibles are much easier to administer and reduce administrative costs, since the insurance plan does not have to reimburse patients for thousands of small bills.

On the supply side, we first summarized the impacts of different payment methods. Table 8.6. shows the impact of each payment method on health care costs and the quantity and quality of services. Empirical studies have found that payment mechanisms have measurable effects on: 1) the modality of medical care provided to patients (e.g., medical versus surgical treatment of angina); 2) the types and amounts of drugs prescribed; 3) the quantity of services provided per visit or per day of hospitalization; 4) the length of stay per hospital admission; 5) the proportion of patients treated on an outpatient versus inpatient basis for a given disease; 6) the labeling of the disease and its severity; and 7) the frequency with which patients are referred to specialists and laboratory tests.

Table 8.6. The Impact of Financial Incentives

Payment Mechanism	Impact on Medical Decisions and Costs
Fee-for-service	Providers favor this method; quantity of service per patient and total supply increase; inflationary; quality may decrease due to overtreatment and overuse of drugs.
Per Case	Improves efficiency of hospital services; establishes uniform financial and clinical information systems; DRG creep; increases admissions somewhat; quality may decrease because of too short of stay per admission and underuse of tests.
Per diem	Less inflationary than fee-for-service; significantly increases length-of-stay ; quality of services per day may decrease.
Capitation	Significantly reduces unnecessary services; improves efficiency; patients maybe undertreated; risk selection by providers.
Global budget	Improves efficiency; most effective in controlling inflationary health costs; quality may decrease; quantity may increase if volume standard is tied to the budget.
Salary	Removes incentives for over- and undertreatment of patients; quantity of output per-hour may decrease; quality of care may decrease; providers self-refer patients to their private practice.

As shown in Table 8.6, no payment method is perfect – each has some positive and negative attributes. However, international studies have found that the magnitude of their positives and negatives are measurably different. We offer several conditional guidelines based on prevalent international experience.

- Decisions on payment method must be considered in a macro-organizational context; they must complement each other.
- Fee-for-service payment method causes rapid health expenditure inflation. Avoid this method unless there are clear and strong reasons that justify it.
- Salary-plus-bonus payment method is superior than salary only. The former can motivate health professionals to increase productivity and improve quality of services.
- In most situations, capitation payment to GPs is the preferred method when there are competing GPs serving the same community. As for specialists, salary-plus-bonus payment is preferred, but that implies the specialists would be employed by organizations such as hospitals or insurance programs.
- For high- and middle-income countries, per admission payment or simplified

DRG payment to hospitals are the preferred methods. For hospitals in low-income countries, global budget is the preferred payment method.

Establishing a reasonable payment level requires the consideration of a complex set of market and regulatory factors. If the level of payment is set too low, the providers can decide that they will not participate, or they may decide to participate but charge additional amounts for patients to pay out-of-pocket, which is called balance billing. If the law forbids balance billing, providers can ask patients for under-the-table payments. On the other hand, if the payment level is set too high, payers will have to bear the burden and pay higher premiums or taxes.

Setting the right payment level is therefore an extremely contentious and sensitive affair. Competitive bidding is an ideal approach for doing this when there are competing providers, but this condition may not exist or providers as a group may not accept the bidding process as a means for setting the payment level. Alternatively, a “reasonable” payment level can be set based on data, but providers will rarely accept one common set of data as being sufficiently objective and unbiased to set a "reasonable" level of payment.

The simple fact is that payers want to keep the payment level as low as possible, while providers want to raise the level as high as possible. Each side can offer many reasons why the payment level is too high or too low. International experience shows that bilateral negotiation may produce the best mutually acceptable results where a committee of payers (those who have to pay the taxes, premiums or out-of-pocket) directly negotiates with a committee of providers (those who receive the money.) If the government intervenes and enters as a key decision-maker or arbitrator, the situation quickly becomes politicized: both sides will exert maximum political pressure for the government to favor their position, and government will then find that it has to use general revenues to subsidize the payment. However, this will distort allocative efficiency.

References

Harshbarger, Melanie. “Note on Physician Compensation and Financial Incentives,” Harvard Business School, 1999.

Kralewski et al. “The Effects of Medical Group Practice and Physician Payment Methods on the Cost of Care” Health Services Research. 35(3): 591-613, 2000.

Oilman BH. “Hospital Response to DRG Refinements: The Impact of Multiple Reimbursement Incentives on Inpatient Length of Stay”, Health Economics 9: 227-294, 2000.

Chapter 9
Organization

1. Introduction

How does the organization of health care provision influence health system performance? More specifically, how can we change that organization to improve health status, satisfaction, and risk protection? This chapter, on the " organization" control knob, tries to answer those questions.

This topic can be seen as composed of two complementary parts. One, macro-organization, encompasses the broad structural features of health care delivery like the public-private division of labor, the kinds of providers available and the degree of decentralization and competition. The other: micro-organization focuses on the incentives, structures and capacities *inside* of provider organizations. This chapter is broadly organized along these lines, dealing first with macro-organization and then with micro-organization issues.

However our analytical approach to some extent blurs this distinction. This is because these two parts of the organization control knob often affect each other and this interaction is important in determining their impact on health system performance. One way in which macro changes influence performance is through their impact on micro level factors. Conversely, how things work at the micro level is an important determinant of the likely effectiveness of macro-organizational change.

Because of this, our general caution that "the devil is in the details" is especially relevant when it comes to using the organization control knob. How any particular organizational reform is designed and carried out will have a significant impact on its actual consequences. The variation in outcomes "within" a given macro-organizational strategy can be as great or greater as the variation "between" approaches at the macro level. Successful use of the organization control knob usually requires attention to detail at both the macro and micro level.

Health reform debates often focus much of their attention on macro-organization issues. These are an important focus of this chapter. But here we will also discuss strategies aimed directly at the micro level, especially those related to the government's role as a health care provider. These include changes in the internal organization of public providers (including "autonomization" or "corporatization"), the use of contracting to provide incentives and techniques to improve performance within the public sector itself (e.g. Quality Management).

First however we have to answer the question of *why* would government want to try to turn this control knob? That is, what are the kinds of performance problems to which changes in organization might be the answer? The short answer is – macro and micro organization matter for how well the health system does in terms of the intermediate criteria such as access, efficiency, quality, financial burden, and cost. And

they matter also for the distributional effects related to these criteria. And these linkages lead to effects on the ultimate performance criteria of the health system.

As we noted in Chapter 4, technical and allocative efficiency, as well as the level and mix of qualities – both clinical and patient amenities – produced by health care services, all influence health status, satisfaction and risk protection. High costs in producing clinical care is likely to mean less physical availability of services for a given budget and lower health status. The wrong mix of services (allocative inefficiency) will do the same. And higher costs will make any given level of risk protection more difficult for a nation to afford. Moreover low quality – or the wrong mix of qualities – will lower utilization, satisfaction, and usually health status. Efficiency, quality, and access problems also have important distributional consequences, in many cases profoundly affecting the poor and disadvantaged with direct consequences for their health, financial protection, and satisfaction.

We argue strongly here that organizational change can influence technical efficiency (unit cost), quality *levels* and the set of available services. Macro changes in the structure of health care delivery can affect the kinds of services available, as well as more subtly the quality of services available and hence acceptability to various groups of patients. Changing incentives on – and inside – organizations can influence efficiency. Altering incentives, capacity and motivation can impact quality. For these reasons organizational change is often a major part of health sector reform efforts.

2. Conceptual Approach

Let us now be more specific. We assert that the organization of health care provision, and hence the organization control knob, matters for health system performance. Recalling our diagnostic framework, this effect works through the intermediate criteria. But how do organizational choices have an impact on our intermediate criteria – access, efficiency, quality – as well as system cost? That is can we trace more explicitly how they affect what is produced and consumed, and at what cost and link those changes in supply and demand to likely effects on the intermediate and ultimate performance criteria?

On the supply side, we know that the output of the health sector depends, first, on the system of production that is employed, and second, on how those technological possibilities are exploited—that is in how production is carried out. How then does organization influence both what is technically possible and actual production practices?

What is technically possible in the health sector depends, on several features of the organization of the system. The first relates to what economists call “economics of scale.” In certain situations, larger size can, mean better performance. A large organization can lower cost by fully utilizing specialized capital equipment (e.g. X-ray machines). It will also have enough volume to employ specialist personnel who can provide better care.

The kinds of services it is possible to produce also depend on the kinds of organizations that exist to do the producing. Large regional centers offer a different mix of service "qualities" than local practitioners' offices. Technical possibilities also are a function of the inputs available at each site: doctors, drugs, functioning vehicles, working laboratory equipment, etc. And these too are influenced by organizational choices.

These supply-side factors also determine important potential effects on the demand side, for example the costs borne by consumers and the quality characteristics preferred by consumers influence the use they make of available services. Large high technology organizations may be perceived by consumers as better for all types of services, even though clinically this may not be the case. Or, more subtly, larger, urban hospitals may embody cultural, physical, and financial barriers to utilization by the poor and rural population, resulting in potential disparities in access.

When we move from what is possible – with a given set of workers, facilities, supplies etc. – to what actually occurs, we come face to face with an important realization. A large part of organizational performance depends on how various "front line workers" (individual care providers like doctors and nurses, vaccinators and health inspectors, laboratory technicians and clerical staff) actually behave. Moreover the incentives, obligations, and opportunities these workers face are the result of decisions that have been made by health sector managers – at the institutional and systems level. For it is managers who buy equipment, organize the work, decide what services to offer, create payment systems, hire personnel, etc. These individual behaviors, summed up to the systems level, also have an important impact on the supply and demand effects of health care organization. And these individual behaviors are strongly influenced by conditions at the micro organization level, where people actually do the work of health care. In what follows, therefore, we have to ask not only what can be done to effect change at the macro-organization level, but also what difference do macro organizational choices make at the micro level? For these are both important parts of the causal chain that connects organizational reforms to the performance we care about.

To explore this connection we want to pose, and answer, six specific questions, which largely explain the effect of any given organizational choice. 1) What are the incentives on provider organizations? What do they have to do to acquire the resources they need to survive and grow? 2) What are the incentives on managers? How are they hired, paid, promoted and held accountable? 3) What tools, capacities and attitudes do managers bring to their role? How are they recruited and trained? How does their experience tend to shape their values and attitudes? 4) What authority do managers have? Can they rebudget funds, hire and fire, contract, set services, and prices make investment decisions? 5) What incentives do workers face? How are they paid, promoted, etc.? And finally, 6) What about worker capacities, attitudes and values? How are they recruited and trained? What is their experience on the job like? For all of these factors, and the way they come together, influence an organization's performance.

We want to stress however that these factors operate in a context shaped by macro level choices. Macro choices (e.g. public vs. private doctors, or integrated vs. specialized

hospitals) determine the kinds of entities that managers get to be the managers of. Hence they shape the questions managers ask and the answers they are likely to come up with. They also affect the systems of production that are available to managers (e.g. are we large enough to take advantage of the gains of large size?) as well as the kinds of providers available to patients.

How then can government "turn" the organization control knob? A crucial distinction needs to be made between the government's role as a direct provider of services, for example through hospitals, clinics, and public health programs delivered by government-owned institutions; and those parts of health care delivery where government is not the provider, but still can significantly influence others.

Where government is the provider, it can act on both the macro and micro level. It can use its direct authority over public provision to change the mix of organizations providing a service. For example, it might restructure its services from hospital outpatient departments to freestanding community health centers. Government sometimes creates whole new types of health care delivery, for example, community health worker programs. It can also use its ability to "exit" (or enter) service delivery, in practice shifting from using the public to the private sector or vice versa. As a service provider, government can also act directly at the micro level – for example by changing systems of management, governance, accountability, etc. within its own hospitals.

Where government is not the provider, it has to act more indirectly. Lets look at several examples. Government can use regulation (e.g. licensing) to influence the number and character of private providers performing a service and this can have an impact on the levels of competition and the kinds of incentives providers confront – as well as on cost and quality. Government also can change whether it pays and who it pays (financing) or how it pays (payment) for certain services. Financing, payment and regulation can also be used to reach into private organizations and affect their micro structure. That is, such tools can be used provide incentives or constraints related to organizational structure, labor relations, staffing levels etc.

Government can also act on the *inputs* into the health care system and in this way influence macro-organizational conditions. More medical graduates will mean more competition among physicians. Allowing some organizations to acquire high technology equipment (e.g. MRI scanners) and not others might mean, for example that the mix of providers of diagnostic services will be altered. Expanding the supply of pharmaceuticals of a certain type can have implications for the "production function" available to providers and hence their costs and quality.

Here again, government can be both a direct producer of inputs as well as use its powers to influence other producers of inputs.

Where government directly produces inputs, it can influence the mix through its own actions. It can open or close medical schools or change product mix in government-owned pharmaceutical companies, for example.

Where government does not directly produce inputs, it can still influence their production and use indirectly. It can grant or withhold import licenses of drugs and equipment, allow or not allow foreigners to practice medicine, use tax policy to stimulate local production, directly license – and hence control – input producers from private medical schools to pharmaceutical companies.

The actions of government to influence input production, both direct and indirect, affect organization in complex ways at both the macro and micro levels. Changes in input supplies affect the production possibilities and production costs of organizations and the qualities that affect consumer behavior as well.

Clearly, the scope of the organization control knob is complex, including strategies which both directly "turn the knob" as well as those that bring about change through settings on other control knobs.

3. Using the Organization Control Knob: Macro and Micro Strategies

Thus far, we have laid out broadly the why and how of the organization control knob. But in practice, governments try to adjust the settings on macro and micro organization in specific ways. Even a cursory reading of the literature on health sector reform will turn up references to efforts to design, implement, and evaluate organization reform efforts in several different areas. We discuss some of the most important of three below.

As with other control knobs, most notably financing and payment, governments have a dual role regarding organization. They both change how private organizations function and, as important direct providers of services, act directly upon the settings as well. Moreover one set of government policies can influence the government in its other roles. In particular government action on macro organization can influence the place of government providers in that structure. Conversely, governments' action on the micro organization of government providers may have second order effects on macro organization. The challenge in finding the right settings for the organization control knob is finding the right mix of macro and micro strategies.

4. Macro Strategies

4.1. Macro Policy Strategy I: Changing the Public-Private Mix in Health Care Provision

Many nations are debating the role of public versus private providers in the health system. Our first point here is that often this issue is miss-posed. International experience suggests that public or private *ownership* – the main basis for the public-private distinction – is often overestimated in importance as a determinant of health care

performance. More often, it is not ownership *per se* that matters, but the specific *details* of how ownership influences the behavior of organizations and those who work within them. Those details – at the micro level – should be the focus of attention. We also need to understand how those details come about and can be influenced at the macro level. There are many examples internationally where positive results can be achieved with either public or private ownership, depending on how it is done. But to see this point it is helpful to begin with the way the issue is conventionally discussed.

Why do public and private sector hospitals and clinics often behave differently - even if there is substantial "within category" variation? The first point of difference is that public sector providers usually rely on the government budget for funds while the private sector depends on customers. (Of course some public entities are market dependant and some private firms rely on legislative largess but we are discussing the "typical" case.) What difference does this difference make? For-profit businesses in competitive markets have to please customers to stay in business. Their desire to maximize profit also gives them a reason to be technically efficient – since lower cost production increases those same profits. On the negative side, however, private, for-profit providers also will be tempted to create what economists call "supplier induced demand" to increase income. Moreover profit seeking private sellers have no reason to serve unprofitable worthy causes – like providing free care for the poor.

In contrast public sector entities have to please those who decide on their budgets. One strategy is to provide great service and have customers lobby the government and/or organize to support the party in power. An alternative to this "service" strategy is a "patronage" strategy. This involves relying on employees and suppliers to be a pressure group on, and political resource for, the government. Since employees are easier to identify, influence and organize than customers, the patronage strategy is by far the more common around the world. The result of such a strategy is often excess staff, high costs, political involvement in hiring and purchasing, and mediocre-to-poor customer service.

There are also public-private differences in the incentives of managers. Private sector managers are typically hired by Boards of Directors that are in turn elected by shareholders. Their compensation and employment often depends on profits. Indeed some managers are themselves owners – and benefit directly from those profits. Public managers, in contrast, are selected either politically or by civil service processes. They are seldom subject to significant up-side rewards or downside penalties based on profitability.

Public managers also generally have less authority than do private managers. Restrictive purchasing and personnel systems were often put in place in the public sector to prevent public managers from using their power for political ends. But these limits also make it difficult for them to do anything – bad or good. For example, public managers can seldom start new programs on their own initiative, or reallocate funds to support new ventures.

Public and private managers also often differ in skills and values. Well-run competitive private organizations try to recruit managers who have the training, temperament, interest, etc. to do such work. In contrast, many public health sector managers are doctors without managerial training who are not especially interested in a management career. Seeing management as "trivial" or "obvious," and a diversion from their "real work," they don't even know what it is they don't know about how to be an effective manager.

We also need to ask how public and private organizations vary at the level of the front-line worker. In particular what is different about the two kinds of organizations' "control systems"; their arrangements for imposing rewards and punishments on workers (not all of which are financial)? When an organization has a strong control system, the consequences to workers of their own actions and decisions, are both predictable and performance dependant (a salesman on commission confronts a very strong control system). In contrast a "weak" system produces unpredictable rewards not linked to performance.

Again, there are typically differences between public and private sectors. Union rules limit the power of supervisors in both spheres. But civil service rules and/or political influence are additional significant constraints on the ability of public managers to reward and punish employees.

Finally there are public-private differences with respect to the motives of both workers and managers. These are produced and reinforced by processes of selective recruitment, peer pressure etc. In some cases the public sector benefits by attracting employees committed to social goals. The private sector, in contrast, is more likely to attract those focused on their own economic advancement. In addition, exactly because the public sector often provides secure if unexciting jobs, it often attracts more than its share of those who are relatively rigid and risk averse. Private business, exactly because it often rewards the aggressive and entrepreneurial, is more likely to attract those who find uncertainty and opportunity more congenial.

We also need to stress that in health care our choices are really more complex than the classic dichotomy suggests. There is often a "third way" in the form of the NGO sector. NGOs often combine incentives to and accountability of managers with a Board of Directors that is not primarily profit oriented. Thus they – at their best – can produce greater "technical efficiency" (including efficiency in the production of quality) than rigid public organizations. Moreover they can recruit staff with different skills and motives than either for-profit business or classic bureaucratic public agencies. This, plus differences in the way they organize services, can lead to differences in the level and mix of quality they produce.

Moreover inside the public sector, we have a host of detailed micro-organizational choices that can change structure and behavior away from the classic model. As we discuss below, we can alter governance, give managers more freedom, lessen political

involvement etc. and still operate within the public sector strictly defined. And these changes may have a significant impact on efficiency, quality, etc.

In summary, classic public sector bureaucratic agencies often exhibit higher cost and less customer responsiveness than the private sector. The private sector, in contrast, is often more flexible and more ruthless. It is not interested in providing unpaid-for services, no matter what their social value. And profit seeking only produces efficiency and quality when market conditions force private sector managers to make such efforts. Of course there are high performance public agencies (from Public Health nurses in Sri Lanka to the U.S. Marine Corps). And below we explore how to foster increased excellence in the public sector. But still, a broad pattern does exist.

How should health reformers choose between using the public and private sectors to deliver services, then? We have no simple answer or recipe. Health system reformers have to bear conflicting considerations in mind and devise pragmatic strategies. If they are going to use the public sector can they institute changes in financing and payment, as well as at the micro-organizational level, that will improve performance? Can these be done at a reasonable cost, for example a cost lower than simply purchasing services from private providers? If they are going to turn to the private sector, can the government exercise the needed regulatory control (as we discuss in the next chapter) to protect consumers and insure effective competition?

Clearly, "the devil is in the details." We will find exactly the same result when we explore the next topic, the impact of variations in the division of tasks among health sector organizations of varied size, scope and character.

4.2. Macro Policy Strategy II: Changing the Provider Mix

What is the effect on performance of the kinds of providers that do the work of the health sector—regardless of whether those providers are public or private? And how can we alter that mix to improve health sector performance? For example, does it matter if hospitals do a great deal of outpatient care (as in China) or almost none (as in Germany)? Is it better to have primary care provided by independent general practitioner (as in the U.K) or by clinics with multiple specialists (as was the case in much of Eastern Europe)? Are programs to control infectious disease best organized separately – or as part of the general primary care system? There are so many issues and alternatives here that it is difficult to review all the relevant material concisely. We will try to do so by first talking about three general characteristics: scale, scope and integration. Then we will discuss the particular organizational problems of providing service in rural areas, and the techniques available to government to change this "setting" on this control knob. Throughout we should keep in mind that the mix of providers can influence performance through both supply and demand side effects.

4.2.1. Scale

The first decision system managers face about provider mix involves issues of scale. We noted above that larger scale can produce lower costs. The extent of these advantages however are different for different health care activities. Hospitals may continue to benefit from lower costs up to say 300-400 beds. Primary care providers get no cost advantages from being very large, for example having many physicians working together in the same facility, although there are some benefits in forming practice groups. Moreover at some point diseconomics of scale appear. That is, costs begin to increase as size increases further. In part this occurs because very large institutions often suffer from a lack of coordination and communication. They can also find it difficult to motivate staff to high levels of effort. Employees feel insignificant and hence unresponsible for actual outcomes. Finally, workers (and patients) have to travel longer distances in a system with fewer and large centers so the costs to society (including patient time) tend to go up.

The advantages and disadvantages of large-scale will thus depend upon each country's particular situation. Densely settled, urban environments provide enough population to support multiple large-scale providers. The same will not be true in rural areas, or even in modest sized cities. Uruguay is not Uzbekistan.

A second consideration in designing a delivery system is that doing more of something makes one better at it. In fact these "practice effects" operate at both the individual, and team level. And clinical skill depends on both the current rate of "throughput" (how many cases are seen per month) and the total cumulative experience providers have had with a particular type of case. Indeed those advantages have become larger in recent years as medical knowledge has expanded and medical equipment and techniques have become more esoteric. However since many acute conditions are not all that common, a large population base is required to provide enough cases to support a highly specialized care unit.

The problem for system managers is that having a few large referral centers implies increased travel time and cost, especially for those outside major metropolitan centers. This means creating barriers that tend to exclude poor and rural people and those in marginalized groups. When a system has only limited treatment capacity, patients with superior economic or social position tend to get better service.

Moreover the dictates of scale can conflict with those of competition (as we discuss in the next section). Without pressure from competitors, regional monopolists may not achieve their potential advantages because they will not work hard enough to do so. There are also political pressures that run counter to regionalization. Providers left out of such a system can be expected to resist such arrangements strenuously. So too will political leaders who want to defend their local facilities.

Similar conflicts can arise even at the primary care level. A choice among individual family doctors, small clinics with 3 to 5 practitioners, and larger clinics with

20 or so doctors, is also a scale issue. And some of the same tradeoffs among clinical quality, service quality and effective availability exist at all levels.

We have no simple advice to offer about what scale of facilities to create. In part it is a question of which performance criteria (health status vs. satisfaction) matters more. The availability of transportation services, the degree of sophistication among rural people, the burden of disease in a country and its overall level of economic development will all influence the consequences of a strategy that leads to more widely dispersed capacity. Politics too will also vary. Yet for countries with limited budgets, taking advantage of scale and practice effects to improve both cost and clinical quality should, we believe, remain a serious consideration.

4.2.2. Scope

A second question about the nature of health sector providers has to do what activities each provider undertakes. To what degree are they diversified (i.e. do lots of different things) versus specialized? At an institutional level this issue arises in the form of decisions about whether society should support specialized institutions like maternity hospitals, or cancer hospitals, versus general hospitals that care for all kinds of patients. At the primary care level it comes in the form of issues like whether to use G.P.'s or a team of a pediatrician, an internal medicine specialist and an obstetrician/gynecologist to provide primary care.

One reason many health care providers are diversified is to take advantage of economics of scale that develop as a result of total volume – rather than as the result of the volume of a specific kind of case. For example a hospital lowers the average cost of X-ray, laboratory and operating room facilities if it uses these to care for more patients – even if it has to care for lots of different kinds of patients to get that volume. Similarly, a small town may not provide enough volume for a specialized obstetrician, but it might be able to support a family doctor who also performs deliveries. A broader scope of practice thus reduces the necessary population base and hence allows for a denser (and more accessible) set of providers. The epidemiological shift to chronic disease is also relevant here since the effective management of chronic conditions requires relatively frequent patient contact. This in turn implies additional reasons to create an accessible system of primary care.

On the other hand, diversification risks losing some of the gains of specialization. Recent industrial experience suggests that highly diversified companies may end up not doing any one thing well. An organization with a narrow product line – in contrast - can attract workers with a particular interest in, and a commitment to a particular activity. The organization can develop an "esprit de corps" based on a relatively clear mission. A well run specialized heart institute or smallpox evaluation program can benefit from such focus in a way that a general hospital or a general primary care system cannot (provided of course that the managerial capability is there to take advantage of the potential gains).

What advice can we offer to health systems planners about the appropriate scope of different providers? For primary care, where access is critical, broad scope – and hence

a need for smaller population base – has much to recommend it. In rural areas, with less dense populations, a broad scope of practice is likely to be the only option at the primary level. On the other hand, the goal of improving clinical quality through an increase of specialization becomes more salient when dealing with complicated or unusual cases. Thus each nation has to make these tradeoffs in light of its own situation – and may make different decisions in different regions or for different services.

4.2.3.Integration

So far we have been discussing the scale and scope of each provider in the health system in isolation. But in a health care "system" there is also always a question about how various providers relate to each other. That is, there are issues about the degree of "integration" – among various organizations and activities. For example how well do primary care doctors relate to specialists or do the doctors at health centers connect to the nurses who make maternal and child health visits? How well connected are diagnostic facilities to treatment facilities, specialized infectious disease programs to the general care system, or school nurses to other providers? Similarly when separate public hospitals are operated by the Ministry of Health, the Social Security fund, the military, etc., is there any integration among these systems?

Integration can be achieved bureaucratically – by having various providers be part of a single system. In theory such arrangements make it possible to rely on hierarchical authority to create and maintain channels of information and coordination. Such systems often rely on primary care doctors to do "gate keeping" and control access to specialists. But gate keeping only works well when the gate keepers don't have any strong incentives to either under or over refer (fee-for-service general practitioners may do too much, salaried practitioners supervised by weak managers and with civil service protection, too little). And in any case, the managerial and administrative demands that must be met to have such a system function effectively are very considerable. It is not easy to make a large, bureaucratic system run efficiently.

In countries where self-referral to tertiary hospitals is possible, institutions with a reputation for higher quality care are often over-burdened. Patients by-pass peripheral centers that they perceive (perhaps correctly) offer lower quality care (because they lack drugs, equipment or well-trained practitioners). Clearly creating an "integrated" system under these conditions requires not only improving quality at the periphery – but also influencing the perceptions of citizens. And do note, it may not be possible to do the latter without first addressing the former!

When the system is not centrally run – the only way to achieve integration is to utilize various voluntary arrangements. Here there are lots of examples around the world. The tumor registry in Aachen, Germany is run by the university teaching hospital. It regularly organizes meetings of local physicians to review cases and to discuss and diffuse new treatment techniques. Similarly WHO has organized groups of obstetricians in various countries in the Caucuses. Participating doctors collect data on maternal and infant mortality and diffuse information about both outcomes and best practices. The medical society in the relatively rural state of Maine in the U.S. has created an N.G.O. to

monitor physician patterns of practice. When it discovers large variations in practice patterns, it assembles a group of experts to work with those doctors who deviate most from the norm. The goal is to convince the outliers to accept standard decision rules and techniques.

Experience suggests that there are limits to the ability of the state to achieve integration by coercive means. For example it is hard to prevent patients from bypassing lower level facilities they don't trust. As we discuss in the next chapter, enforcement is seldom effective when those subject to the regulations don't recognize them as legitimate. This is especially so when it comes to problems like improving clinical practice. Integration is a desirable goal but it needs to be fostered rather than coerced. Hence various means (incentives, certification, social marketing, etc) that enlist patients and providers positively in efforts at integration are often the most effective option.

4.2.4.Providing Services to Rural Areas

In many low (and even middle) income countries a particular "provider mix" problem involves providing services in rural areas. In many countries, the standard approach has been to construct a network of health centers (with a few inpatient beds and physician staff) and subcenters (with non-physician staff and no beds) and rely on the Ministry of Health to operate these. Yet in many nations the same complaints are heard. Doctors don't want to go to remote areas, and if assigned don't fulfill their responsibilities. Drugs are lacking, equipment is not maintained, and the resulting poor quality discourages use by even the poor. Even in areas where government services appear to be available, populations may use a variety of traditional healers and not formally qualified providers as sources of care, who they pay out-of-pocket.

In countries where physicians in the public sector supplement their incomes with private practice, poor rural areas are unattractive because they offer few such opportunities. In addition such areas offer little in the way of educational, social or economic opportunities for physicians and their families and may even be unsafe – especially for young unmarried women doctors. Moreover national systems based on uniform salaries limit the use of either incentives or coercion. Supervisors seldom visit remote posts, don't have many coercive tools to use, and can't pay doctors more to locate there. In such a situation, simply building more facilities – that will go understaffed, under-equipped and underutilized – is seldom an effective answer. There are some apparent exceptions where the Ministry of Health does seem to be able to provide service in rural areas, like Sri Lanka and Cuba. But these are indeed special cases: small islands, with relatively dense populations and a high degree of ideological motivation in the public sector.

Some countries have found ways to provide incentives to rural doctors. Some Indian states give preference for admission to specialty medical training to those who have served two years in under-served areas. Countries as diverse as Canada (for Arctic communities) and Uzbekistan have created differential pay levels to attract doctors to rural areas (in Uzbekistan by paying one doctor more than one salary!) In rural areas of

Kerala, in south India, local governments are allowed to supplement the statewide salary scale to attract doctors.

Changing organizational forms can also help. In Andhra Pradesh, India the state has contracted with NGOs for the operation of some health centers. These contracts give NGOs substantial management flexibility so that they can draw on staff with a particular social commitment. In aboriginal communities of Australia, public funds that used to operate state run medical services are now sometimes turned over to local NGO's (supervised by community boards) to operate locally controlled medical services.

Another approach is to recognize that a uniform system of services may not be appropriate in a socially and geographically non-uniform country. Maybe not all centers should have doctors. Instead less trained personnel, supplemented by weekly or bi-weekly visits by well-equipped mobile vans may be an alternative (one being used in some remote areas of Mexico). Australia supplements care in remote areas (offered by a resident nurse) with the well-known "Flying Doctor" service, which is both an evacuation system and a mechanism for bringing in doctors to run periodic clinics in remote areas. In sum, changing provider mix may be critical for improving services in remote areas – and for in turn increasing both health status and satisfaction.

We also note that many of these issues are equally relevant for disadvantaged urban populations. Many governments have established urban primary health care delivery systems that are similar in design to rural services and often suffer from similar problems of low utilization, while poor urban populations pay out-of-pocket for private providers.

4.2.5. Influencing provider mix

In directly operated systems, provider mix can be charged by the use of hierarchical authority. However, as we noted in Chapter 4, such changes can be extremely difficult politically. This is especially so if some areas are to lose services (or employment) as a result of the proposed reforms. Moreover unified systems don't always find it easy to get patients to cooperate with the government's view of "rational" use. In particular almost everywhere patients can, and do, go to private practitioners – including those the authorities view as less than fully qualified – when that option seems best to them.

One strategy often used by governments is the creation of new systems of health care delivery. Community-based health workers may be the best known example. Vertical immunization and family planning programs are another example. These strategies have been attractive because they use lower salaried or volunteer workers with less training and they offer the promise of mobilizing community resources.

Our advice to readers is to be careful about assuming that changing provider mix will lead to cost savings. The services of one community worker may be relatively low cost, but developing and supporting thousands of them may not be. Strategies that may be attractive because they appear to be "low cost" may not always be so. Many community

health worker programs failed because – due to budget limits – they did not provide adequate training, supervision, and supplies.

A variety of regulatory and licensing techniques are available. to control the services offered by, and even the organization of, the private sector. Recent attempts in Korea and Japan to "disintegrate" physicians from their traditional pharmacy role is one example. Another is the use of so called "certificate-of-need" requirements that create a licensing scheme for new services. We discuss these ideas further in the next chapter, on regulation. Here we just want to preview that discussion by noting one of its key conclusions – namely the effect of such efforts depends critically on how they are implemented. And regulations that run counter to the strong fiscal interests of the private sector are often difficult to enforce.

Finally, the financing and payment systems are also powerful tools for influencing provider mix. In the U.S. the government insurance program for the elderly (Medicare) has concentrated certain services in regional "centers of excellence" by refusing to pay for those performed elsewhere. It also created whole new classes of providers (e.g. free standing kidney dialysis centers and home care agencies) when it decided to pay certain kinds of providers for certain kinds of services. Similarly some regional Polish health insurance funds have fostered a shift in outpatient care from polyclinics to family doctors by preferentially contracting with the latter. Thus effectively turning the macro-organization control knob, with regard to provider mix, may well require mutually supportive and complementary efforts to turn other control knobs as well.

4.3. Macro Policy Strategy III: Fostering Centralization or Decentralization

Our third and fourth major macro-organizational issues apply to the public sector and the private sector respectively. In this section we deal with decentralization in the public sector, and in the next with competition in the private sector.

The argument for decentralization in the public sector again links macro-organizational features of the system to incentives and motivation at the managerial level. The underlying argument is that centralized national systems cannot (or at least often do not) provide effective supervision or incentives at the local level – especially in large and diverse countries. Unhappy citizens face too many bureaucratic obstacles in making their complaints heard. With many facilities to monitor, managers at the center cannot keep track of – or even have much reason to be interested in – performance deficits at the local level.

Decentralization advocates contend that shortening chains of control and accountability will improve service delivery. Poor performance will be more visible and the pressure managers feel for doing something about it more acute. Patronage and corruption will be more visible, and the price they extract (in the form of lessened performance) will be less acceptable. Citizens will have someone to go to with complaints, someone who will have reason (electoral success) to respond to their concerns. Managers connected to local or regional political institutions will be held accountable by those institutions for system performance. All this – it is claimed – will

lead to lower costs (improved technical efficiency), as well as higher quality service for the available resources.

These same political processes, it is alleged, will make it possible for variations in local circumstances and preferences to be incorporated into policy. While national bureaucracies don't have to ignore local conditions, decentralization advocates argue that in practice they often do just that. They argue that local control, in contrast, will lead to changes in what is produced. Some of these changes will improve allocative efficiency and hence performance with regard to health status goals (e.g. more money will be spent on malaria control in malaria prone areas). Other changes will improve performance with regard to customer satisfaction – by producing a mix of outputs, and the kind of service that respond to local preferences.

This account however oversimplifies the options available to a would-be decentralizer. Decentralization does not have to involve a shift in control to local political entities (often called "devolution"). It might instead result in stronger local or regional authority within a national system ("deconcentration"). And not all decentralization involves all aspects of managerial authority. Some decisions may be reserved to the center; so that what is decentralized (what Bossert calls the receiving organization's "decision space") can vary substantially from case to case.

Decentralizating responsibility for financing raises further complexities. Richer regions can finance the same services with lower taxes (or more service with the same taxes) than their poorer counter-parts. So unless some sort of inter-regional redistributive mechanism is established, poor regions can be victimized by decentralization. On the other hand where sufficiently strong redistributive efforts are made, poor regions can actually gain from decentralization! As evidence from Chile and Columbia suggests, formula-based redistribution may treat poor regions more generously than prior, politically-driven, budget processes.

On the other hand, in countries where local political processes are themselves non-transparent, elite dominated and patronage ridden, the hoped-for accountability and efficiency advantages of decentralization are not likely to materialize. Moreover in countries with significant inter-community conflict, marginalized groups may find themselves worse off with local decision making.

As may seem obvious, capacity building at the local level can be critical to making decentralization a success. On the other hand, diffusing technical skills and information to local leaders will not necessarily overcome a deficit in what Putnam has called "social capital." This is the willingness of citizens to cooperate in their joint interest in ways that involve a certain amount of trust and reciprocal respect. Double-entry bookkeeping can be taught much more easily than the habits of mind required for mature democratic citizenship. In Kerala, in south India, where literacy is high, and where there already is substantial "social capital" in the form of local cooperation, informed observers credit massive training effort of local leaders with greatly facilitating the successful decentralization of health sector budget authority to the local level.

It can be a mistake however to delay implementing decentralization until all the needed training is "finished." That day may never come. Moreover the pressure of dealing with the problems of implementing a decentralization program can produce "teachable moments" that increase the receptivity of local and regional officials to new skills and concepts.

Decentralization in the health sector is sometimes undertaken for reasons other than that of improving health sector performance, e.g. as a way of invigorating local governments by giving them more meaningful responsibilities. It can also be done simply to shift some financial responsibilities away from a hard-pressed national government. In such cases, a decline in health sector performance may be a price society is willing to pay for achieving these other objectives.

Once an effective program of decentralization has been undertaken, the role of the central bureaucracy changes. It now becomes necessary to advise instead of control, train instead of command. Monitoring becomes even more important. Otherwise the effect of the new arrangements will not be known. And various regulatory or incentive devices may have to be employed to push local policy in directions consistent with national goals. For example, cash transfers to subordinate levels of government can be adjusted upward or downward based on the degree of achievement of national objectives. (In India where most health matters are decentralized to the states, the national government has provided substantial cash prizes to the states with the best population control performance). Needless to say, many central government units find the shift in their role – from control and command to information, regulation and negotiation – quite difficult.

What advice do we have for health systems reformers contemplating decentralization? Our first point is that as much as any area of health reform, decentralization illustrates our general contention that implementation matters (e.g. Will there be enough redistribution to improve the lot of poor regions or so little – as in Bosnia – so as to worsen their situation?). Decentralization is an arena full of unanticipated complexities. There is a large and interdependent set of choices to be made with regard to exactly *what* authority is decentralized. And these need to be made in a consistent and coherent way if the potential benefits of decentralization are to be realized. If we expect improved efficiency, or outputs more closely tailored to local circumstances, local decision-makers need to have authority over personnel and/or budgets if they are to have any hope of producing such results.

Our second point is that the balance between the advantages and disadvantages of decentralization will depend on local conditions. How well developed are local political administrative and legal institutions? How much technical and administrative capacity is there at the periphery? All of these considerations should enter into the design of any decentralization program.

The importance of context suggests that health reformers need to think about decentralization initiatives that are both varied and iterative. Not all local areas will have

equal need or capacity and thus should *not* necessarily all be treated alike. As local capacity expands and experience develops, more authority can be transferred. Even in the U.S., where there is joint federal-state responsibility for enforcing certain environmental laws the extent of decentralization varies – depending on both each states' desires and on their demonstrated capacity.

4.4. Macro Policy Strategy IV: Fostering Competition In Private Markets

Just as the degree of decentralization is a critical macro-organizational choice in the public sector, the degree of competition is likely to be very influential in shaping the performance of privately provided health care. As noted earlier in this chapter, competition matters because for-profit sellers only keep down costs and prices and respond to customers because competition forces them to do so. Monopolists too do make higher profits with lower costs, but the effort required to keep costs low is considerable. And if profits are available without effort – why work so hard? As the English economist John Hicks once said, "The best of all monopoly profits is a quiet life." Hence cost control is often dependent on competitive pressures.

The appropriate degree of competition in health care is complicated by several unique features of the sector. First, when providers are paid fee-for-service, increased competition can lead them to exploit patient ignorance, create additional "supplier-induced" demand, and increase the volume of inappropriate care. And this needs to be taken into account in making macro-organizational choices.

The fact that in most health care markets people do not travel that far for care also leads to tension between competition and economics of scale in the hospital sector. Only large cities can support a sufficient number of hospitals and specialists, to create the preconditions for a competitive market. Hence the policies used to produce economics of scale (e.g. licensing to limit new facilities) also often diminish competition.

Yet another complexity is that many forms of quality regulation in health care – including training and licensing requirements – also limit competition. Professional societies that are supposed to play a role in improving quality can likewise act both formally (through pricing agreements) and informally (through social pressure) to facilitate anti-competitive behavior. In some cases, indeed, governments have even given medical societies the ability to restrict the entry of new practitioners, a move with obvious anti-competitive implications.

Given all these factors, around the world many health care markets are only "imperfectly" competitive. They are what economists call "oligopolies." They have only a few sellers – more than the monopolistic one and less than the competitive many. As a result the competitors recognize their interdependence – one doctor or hospital knows that if they raise rates, or introduce a new service, or modernize their facilities, their competitors are likely to be provoked, or forced, to do likewise.

To be fair we should also note that there are some who argue that too much competition is not desirable. Some countries have relied on providers to take the profits they earn from the care of some patients and use them to finance research or teaching or the care of other patients. Competition – by squeezing down profits – diminishes the funds available for such cross-subsidy.

As in other questions of macro-organization, the impact of competition on performance plays itself out at the micro level. Managers of firms that might not survive due to competition have a powerful incentive to be hard working, and risk taking. For if the company fails, they could be out of a job entirely. Indeed there is some evidence that firms managed by their owners are more successful – and take bigger risks – exactly because managers have so much to lose, and gain, from variations in their company's success. (This is why so many for-profit firms reward managers with stock options and other kinds of ownership positions that increase a manager's financial interest in the corporation's profitability.)

Competition, and the risk of failure, creates incentives for workers as well as managers. As the English essayist Samuel Johnson once wrote, "The prospect of hanging serves wonderfully to concentrate the mind." Workers who know that a failure to adapt to the marketplace could result in the loss of their jobs may be more willing to accept changes designed to increase the firm's competitiveness (like new production process).

Given these linkages, competition can be a powerful motivator to improve organizational performance. Properly regulated, (e.g. to counteract the agency problem) it is a potentially useful tool for health sector reformers. Competitive businesses will only do what you pay them to do. They are responsive to market demand, conscious of controlling costs and try to please customers exactly because that is what they have to do to survive. Many poor citizens in poor countries pay substantial sums out-of-pocket for private health care exactly because they are seeking these advantages, and are trying to avoid the poor service, inconsiderate treatment and lack of responsiveness in the public sector.

We begin this last section with the presumption that competition was a tool that applied primarily to the private sector. But in truth that is an oversimplification. We have noted several ways government could encourage competition through use of the financing, payment, and regulation control knobs. But government providers also can be instrumental in policies to increase competition. Most obviously, governments can directly create competition for private providers by establishing or improving existing public providers. Malaysia has been d as an example of a country where public providers' low costs and high quality restrain private providers in the market. Public sector entities too can benefit from the incentives that competition can generate. For example – suppose public sector workers and managers were put at risk if their organization did not perform satisfactorily! Conversely suppose they participated in some of the "upside" gains of marketplace success in the form of increased compensation based on profitability?

4.5. Macro Policy Strategy V: Using Contracting

In a previous chapter we discussed payment systems—where central authorities (the Ministry of Health or a social insurance fund, for example) establish a price schedule for paying doctors, hospitals and other providers. Typically such systems have a high degree of formality and regularity so that all qualified providers in a given class are paid according to a uniform system. Yet while this is the most common approach to paying for health care, there is another strategic alternative. That alternative involves *contracting* with various providers for services in a way which is *not* necessarily uniform. This allows a nation to provide significant incentives even where purely competitive markets are not present.

A contract can provide for a much more detailed and extensive set of incentives then a payment system. It thus combines some of the coercive features of a regulatory scheme with the incentive effects of a payment system. As such it may be particularly appropriate when reform leads to hospitals with increased autonomy and independence. Exactly because autonomized providers are subject a diminished degree of hierarchical supervision, the incentives provided by contractual relationships can help fill the gap and move providers' behavior in the direction of social goals.

To begin with the basics, a contract is nothing more than an agreement between two parties. The buyer agrees to give the seller money, and the seller in return agrees to provide certain goods and services. The contract specifies the terms of that agreement: what the seller provides, what the buyer pays, delivery and payment dates, the time period of the agreement, renewal provisions (if any), penalties for non-performance and often an appeals process if there is a disagreement on the meaning of, or the fulfillment of, the contract.

The relationship created by a contract may be a simple arms-length one (you deliver so many tons of wheat to such and such a place and we send you so much money on delivery). Or it may be a more complex one. The seller for example may agree to provide certain data, or create certain internal quality control processes. The buyer may provide working capital in advance of delivery or underwrite certain debts or assume certain risks.

Many successful contractual relationships in the business world rely on a high degree of reciprocal knowledge and trust between buyers and sellers. The two parties exchange information and work together in mutually beneficial ways, in the context of an ongoing relationship. Sophisticated managers realize that the narrowly defined terms of a written contract often cannot cover all the unanticipated contingencies that arise. And dealing with such unforeseen developments requires a certain degree of reciprocity, based on a commitment on the part of each party to take the other's interests seriously.

It is also worth noting that the processes of reaching a contractual agreement can vary substantially. Sometimes there is a formal competitive "tender offer" process – where the low bidder wins. Sometimes bids are evaluated on quality as well as price, and

a “score” assigned. Sometimes potential bidders have to submit proof of their qualifications, and only those deemed qualified are allowed to bid. Sometimes negotiations are done “sole source” –or perhaps with two or three potential contractors in an informal competition. We would argue that highly formalized procedures work best with relatively standardized (or at least easy-to-describe) outputs. On the other hand, the more discretionary processes involve greater risk of favoritism or corruption. Hence a device like an outside review panel, to preserve both the appearance and reality of fairness, may be required.

Contracting can go beyond providing external incentives and can also be a way to influence the internal structure of providers. That is, contracts can be written to require contracting hospitals to establish certain processes, or hire managers with certain qualifications, or provide certain staffing levels, etc. For example the hospital might agree to set up certain peer review processes to look at all deaths that occur among its patients, or to create an independent body to ensure that all research projects on its patients met certain standards. Such requirements build on the insight that an organizations in many ways can be viewed as a political system – rather like the political systems we discussed in Chapter 5. By that logic, creating or strengthening internal groups that advocate for certain points of view is a way to change the behavior of the organizations as a whole. Moreover creating certain processes can lead to desirable incentives at the individual level – if only the desire by doctors to avoid embarrassment for poor decision-making in front of their peers.

In the context of a system of autonomized or independent hospitals, selective contracting can be a way to achieve important goals. For example contracting can be a means to concentrate certain services in regional “centers of excellence.” On the other hand, this power is not unlimited since the success of a contracting approach generally depends on the degree of competition among those being contracted with. It is much harder to negotiate successfully with a monopolist – who knows the buyer has no where else to go.

In general, the larger the buyer, the more they can demand favorable terms from the seller – up to a point. However when the buyer is the sole customer, the dynamics change. For then if the buyer is too tough, the seller might go bankrupt. And faced with such a prospect, sellers will mobilize politically to protect themselves. When both buyers and sellers are monopolists (what economists call “bi-lateral monopoly”) the outcome is particularly difficult to predict. For then buyer the needs the seller (e.g. the only hospital in a region) as much as the seller needs the buyer (the only purchaser of services).

In sum, contracting can allow a purchaser of care increased flexibility, and precision of control. Different hospitals, in different circumstances, can be paid differently. Hospitals and doctors can be required to file detailed reports. Structures, processes and personnel can be influenced. And quite complex payment arrangements can be created. For example, the Ministry could agree to pay a hospital a fee to cover certain fixed costs plus a volume based fee to cover variable costs. This creates both productivity incentives and offers the hospital protection against financial risks, if volume

declines. And contracting allows such a system to be tailored to different hospitals' different financial situations.

Whenever the "deliverables" under the contract are reasonably complex, the processes of negotiating the detailed contract, and administering it once it is in place, are very important to outcomes. Sometimes an agency inappropriately focuses on *who* wins the contract. Yet what they win (the terms of the contract), and how the winner is dealt with over time, is often what determines the performance that the contract actually produces.

Given these realities, countries that want to go the contracting route need to consider carefully what internal capabilities they will need to do so successfully. Do they have the expertise to develop specifications, make contracting decisions, write actual contracts and oversee compliance with terms they agree to? Contracting does not end the role of the Ministry of Health, for example. But it does substantially change the nature of that role (and the associated needs for skills and personnel).

Contracting does have its down side. All this specification writing, bid reviewing, negotiating and monitoring raises administrative costs above those associated with using a simple, uniform payment schedule. The very discretion involved in the negotiation process also creates the same risks of patronage and corruption that all managerial power carries in its train. Thus the applicability of this micro-organizational approach depends in part on the sophistication and probity of the relevant administrative agencies. Moreover unless the managers of the organizations that face these incentives have the authority, motivation and skills to respond, much of the potential good contracting can accomplish can easily be dissipated.

5. Micro Strategies

The previous discussion of macro strategies raises a number of challenging opportunities for health reform. It also leads us to confront some provocative questions about how to bring about real performance improvements in the public sector. We have argued thus far that the impact of macro-organizational change depends how these are played out at the micro level. But it is also the case that – especially where government operates health care providers – it can directly produce micro-organizational changes as part of health sector reform – the question to which we now turn.

5.1. Micro Policy Strategy I: Corporatization and Autonomization

The first micro-organizational intervention we want to discuss is prompted by the interest around the world in recent years in changing the organization of public hospitals through various devices sometimes referred to as "corporatization" or "autonomization." The underlying idea is quite simple. The traditional public sector – as we have discussed – has too often been characterized by poor performance, poor management, weak incentives to workers, patronage etc. The thought is that if some of the organizational

features that have produced that poor performance could be changed, that performance might be improved.

This view has lead to the creation of a whole range of new organizational forms that seek to import some aspects of private sector management and competitive markets into the public sector. Managers are given more authority and in turn are made accountable to some new governance structure – usually a Board of Directors of some sort. The organization's financing may change – with it allowed (or required) to get more revenue from customers and more capital from the private market. The system can be freed partially or wholly from government budget, personnel and purchasing rules. And the hospital can be paid in various new ways – based on patient volume, or incentive contracts etc. These new organizations are sometimes referred to as "autonomized" or "corporatized" based on their authority and/or their structure. They are also called "quasi-public" or "para-statal" or other terms used to distinguish the fact that they are both in the public sector yet are treated differently from normal state run bureaucracies.

Efforts at hospital reorganization have been quite varied - which is perhaps not surprising given how many different "design decisions" there are to be made. Before we access the impact of such decisions we need to outline the alternatives – which we can do in terms of the critical micro-organizational variables we identified in the beginning of this chapter.

- *Incentives on the organization*: The most extensive reforms have sought to create strong incentives by creating a fiscal situation close to that of an private company with revenue dependant on customer purchases and hospitals able to reinvest profits and acquire capital by borrowing from private sources. Less extensive reforms have relied on incentive contracting and/or so-called "quasi-markets" where public buyers purchase in bulk from public sellers. The most mild reforms have relied on performance based budgeting where some (modest) percent of revenues varies with volume or some other performance measure.

- *Incentives on managers*: The most extensive efforts confer power to hire, fire and compensate managers to a Board of Directors that in turn is insulated from political pressures. This can create great incentives on managers to perform well – depending on how the Board behaves. More modest reforms allow political processes to influence managerial appointments (directly – as in Vienna where the city council appoints managers, or indirectly). Also terms and conditions of employment may be subject to civil services rules, limiting the discretion of those hiring managers.

- *The authority of managers*: The "decision space" of managers, like that of local officials under decentralization, is quite complex. Potential areas of authority include personnel, pay, pricing, purchasing, production etc. More aggressive reforms give managers more authority over more dimensions, although unions often lobby to limit managers' discretion on personnel matters.

- *Managerial skills/attitudes/values:* Much of the responsibility for this has been delegated – perhaps unthinkingly – to the new accountability structures. In relatively few places has there been much thought about the "social engineering" needed to produce the combination of entrepreneurship and social responsibility among managers that reformers would deem appropriate. A few countries (e.g. Hungary) have made major efforts to train new managers and require such training for all new appointees.

- *Worker incentives:* Approaches here vary – depending on the decision space given to managers. If worker incentives are left in manager's hands, they are able to hire, fire, alter pay scales, etc. On the other hand, sometimes these powers are not relinquished by the center, and the personnel system and the incentives it creates remains under central control.

- *Worker skills, attitudes and capacities:* Reformers typically leave these factors to the efforts of the new managers. In some cases, new training programs are put in place – both for in-service employees and new professionals – designed to orient them to the new situation.

Changes made in these various micro-level factors do influence each other. People often choose to work at a job that offers the kinds of rewards they value. University graduates, who want to become rich, become entrepreneurs, not academics. Hence an organization that remains subject to seniority and/or patronage hiring will continue to attract workers who find such an environment comfortable. On the other hand if managers create a new environment and spirit, that can influence who seeks to work with them. And once part of an organization, leadership and education can further shape employees' views and values.

What kinds of hospital restructuring have been more or less successful? While the international evidence is scanty, the little that there is is consistent with theoretical expectations. Piecemeal changes, which leave key aspects of the structure unreformed, are not likely to produce desirable results. Increasing incentives on the organization, without giving managers the authority and skills to respond is not a promising approach. Similarly giving managers authority and subjecting them to market incentives – without any external accountability – is likely to turn them into profit seeking entrepreneurs who slight social values. The autonomized Chinese leprosy hospital, that started a turtle farm when its primary clientele became less numerous, is such an example.

This leads to a second point. Since markets are imperfect, quasi-public or corporatized hospitals may be in as much need of external planning and regulatory controls as private hospitals. Markets and regulation are complements not alternatives as we argue in the next chapter. Indeed the greater the market pressures, and the more autonomized the quasi-public hospitals become, the more they will act like for-profit entities.

Third, the most important changes are often the ones most difficult politically. In-so-far-as the political system does use the public sector for patronage purposes, or insofar as workers are politically powerful, it may be very difficult to get critical personnel decisions turned over to managers. Similarly if the public sector wants to rely on cross-subsidy to fund certain services (as an alternative to taxation) it may be reluctant to have health services be subject to the test of market viability. The political leadership may not want sensible pricing, unprofitable services abandoned, etc. And it may be equally reluctant to explicitly pay for (i.e. subsidize) unprofitable activities.

Those points taken together mean that reformers contemplating autonomization need to think carefully about how such changes will be packaged, and how they will relate to other aspects of health sector reform. Micro-organizational changes need to be quite significant in scope and provide consistent answers to the six key questions with which we began this chapter. In addition reform will often require complementary changes in the financing, payment and regulation control knobs. For example, will the payment system be adjusted to create appropriate institutional incentives? Will the right people get on the new hospital Boards, and will the hospitals be leading appropriately trained and motivated managers? A program of hospital reorganization is not like a strong anti-biotic that—taken by itself, will, in a short time—knock out a bacterial infection and "cure" the condition. It is more like a course of physical therapy, hard tough painful work that needs to be done day by day, over a substantial period of time, and joined to other therapeutic interventions, if it is to do any good.

5.2. Micro Policy Strategy II: Contracting Out

Previously we discussed contracts between fiscal intermediaries and care providers. But a closely related (yet distinct) micro-organizational possibility involves contractual arrangements between care providers (e.g. hospitals) and private sector firms for the provision of certain services. This is increasingly common around the world in the case of hotel (non-clinical) services like food, housekeeping, laundry, etc. But it is also being used in clinical areas (e.g. contracts to perform certain laboratory tests, or to run specialized services like hospice care).

The lure of contracting out is first that specialized suppliers can take advantage of economics of scale in a particular function (e.g. laundry) that an individual hospital cannot achieve. In addition such suppliers can focus all their organizational energy on an activity (like laundry or housekeeping) that might be ignored within a medically oriented hospital. Moreover the hope is that the need to win contracts will provide strong incentives on the contracting organization. In countries where the public sector is inefficient and difficult to reform, contracting out may be especially attractive.

The critical issues here turn on the capacities of the buyer and the capacities and competitive circumstances of sellers. Thus contracting out is most attractive when a well-developed set of competitive suppliers is available. Contracting with a monopolist (even if it is a newly created monopoly supplier as occurred in Malaysia) is not likely to produce the hoped for cost and service gains.

To contract successfully, buyers at the hospital level need the same skills that contracting with hospitals requires at the Ministerial or national level. The hospitals need to be able to write specifications, evaluate bids, draft contracts, monitor performance, etc. Moreover unless the hospitals themselves are under pressure to improve their performance, they will have no more incentive to do contracting well than they formally had to do well at direct service delivery. Thus giving hospital managers the authority to move in this direction, has to be coupled with increased performance-based accountability for those managers, as well as with increased incentives on the organization as a whole.

5.3. Micro Policy Strategy III: Improving Public Sector Performance

We noted earlier, in discussing autonomization, that an alternative to such structural reform – at least in theory – is improving performance within the existing public sector. Indeed the existence of some well performing public organizations suggests that this ought to be possible.

The analysis we offered at the beginning of this chapter – of the determinants of organizational performance – suggests some of the steps needed to get better performance. Some of these are almost too obvious to mention. Adequate budgets, sensible (not political) decisions about where to locate capacity, rational priority setting, etc. Others emerge from our micro-organizational framework.

- *Organizational incentives* that confront public sector bureaucracies. Tie funding to performance in some way. For example, vary budgets based on the quality and quantity of services provided. Alternatively, provide incentives (e.g. prizes) to regions or agencies that do better. Until the organization as a whole has a reason to do better, change will be slow in coming.

- *Managerial incentives:* Select, promote and compensate managers in the public sector base on their skill and performance as managers. Simultaneously establish systems of managerial supervision that create real accountability. Limit politics and patronage, so that, for example, politically well connected doctors are not appointed to important posts who know little about, and care little about, management.

- *Managerial skills and attitudes* also need to be influenced by establishing appropriate appointment and incentive systems. Formal management training can be made a pre-requisite for certain posts. A country can develop real career tracks, give managerial roles prestige and value and encourage those who want to take responsibility for results to pursue a managerial career.

- *Managerial Capacities:*Training both before and after appointment is a key tool. In addition to formal skills, managers also need to be selected for their "natural" abilities. Are they interested in, and insightful about human behavior? Are they outgoing and

communicate well? Are they mature and self-disciplined in their interactions—not only able to accept criticism but actually eager for constructive advice?

- *Managerial authority:* Authority needs to be redistributed to let managers make real decisions. They need to have a significant "decision space," even as we hold them accountable for the consequences of their decisions (see above). Budgeting, purchasing, pricing, personnel, production methods, etc. are all areas where change can be made.

- *Worker incentives:* The personnel system for workers too needs to be changed to reward performance. Such a system in turn must be administered fairly to be credible with the employees. Political influence and patronage too must be limited

- *Worker capacities and attitudes:* What is inside the heads of workers also needs to change. To get new levels of skill and enthusiasm, traditional recruitment patterns, training systems, and personnel practices need improvement.

Clearly such changes would significantly transform many public sector agencies. Professionalizing management; putting aside politics and patronage; providing performance-based incentives – none of this is easy to accomplish. However, countries that give managers little authority, allocate budgets based on past history, rely on patronage etc., should not be surprised if their public sector organizations continue to perform badly!

We do recognize that the politics of changing such a system can be difficult. The unions will resist. The political parties (at local, regional, and national levels) are likely to protest and obstruct. Patients may be opposed, fearing cuts in services, a loss of cross-subsidy or a rise in user fees. The doctors too will often not cooperate – preferring instead to remain unmanaged. That is why strategies like corporatization may seem more attractive – and perhaps more sustainable politically.

Structural changes, however, are not self-implementing. Giving managers authority does not guarantee they will have the skills or the commitment to engineer real service improvements. Since reward systems for workers are always imperfect, change requires leadership, a capacity to motivate and energize. Staff need to feel that the work they do (individually and together) is important and worth doing well. Leaders need to find ways to harness basic commitments among the workforce and connect these positively to the changes they propose. This is not an easy task and finding those interested in and capable of doing it is equally difficult.

In recent years there has been a whole movement into health care of ideas variously called "Total Quality Management" or "Continuous Quality Improvement." Properly understood, the critical components of these ideas are fully compatible with the framework we have been using. Various "TQM-CQI" experts all advocate performance measurement, managerial accountability, worker involvement, a focus on customer service, leadership from the top etc, etc. They also all focus on careful redesign of

systems of production to lower cost and raise quality. A key insight they offer is that doing better on quality cannot come from coercion or enforcement. It has to emerge from a system that makes it easy to do things well and hard to do things badly. When the jargon is stripped away, the so-called "quality movement" turns out to be all about doing basic management tasks better. Targeted efforts in the health sector – to improve drug ordering or inventory management often are based on this sort of approach.

There is now a huge literature on such quality improvement methods that we cannot recapitulate here. Our point however is that such programs are not a substitute for good management. On the contrary, experience shows that these new approaches have their biggest positive impact in the context of already well managed organizations. For it is exactly the existing cadre of managers who have to implement any new ideas. And yet again, implementation is critical to what actually happens! Thus health sector reformers cannot look to such ideas as a way to save them from, or allow them to avoid, the difficult work of reform of the public sector, if they want that sector to actually perform better.

5.4. Micro Policy Strategy IV: Altering the Distribution of Inputs

Our last topic involves direct and indirect government efforts to alter the macro-organizational structure of the system through controlling the distribution of inputs. In fact, almost every nation follows a variety of policies designed to control the distribution of inputs into the health care sector. As we discuss in the next chapter – on regulation – of much of this control is quality oriented – for example, medical licensing or the regulation of drug manufacturing, etc. These efforts are often focused on limiting supply to counteract the potential impact of supplier-induced demand. But measures undertaken for such purposes also affect the micro-organization of the system – as do policies aimed directly at such ends.

5.4.1. Personnel

In many nations, much of the system of education of health sector personnel is directly operated by government – medical schools, nursing schools, etc. Changing the number and mix of personnel trained in such institutions can have a variety of effects on the structure of the health care system. First, sometimes such training efforts are necessary to complement other aspects of organizational reform. For example, a change to a system of family medicine doctors will only work if medical education is adjusted accordingly. Similarly a decision to make expanded use of non-physician providers to "triage" patients in primary care settings also depends on having sufficient supply of such providers. In Australia, an effort to expand the number of doctors in rural areas lead to a reallocation of medical school places to a new institution focused on preparing physicians for such roles.

Second, the number of personnel are likely to have implications for market behavior. Training more physicians will lead to more doctors in any given market area. This might well increase price competition (as well as supplier induced demand for inappropriate care).

In contrast, a substantial excess supply, especially in a largely state-operated system, can lead to enormous political pressures for overstaffing. This in turn can lead to quite inefficient systems of production in the public sector.

Even where government does not directly operate all the training institutions, there is much it can do to affect the flow of new providers into the system. In many countries private medical and nursing schools are licensed by the government – which can use its authority to alter the mix of those trained. In addition where governments want to expand private sector training they can offer a variety of financial incentives and support to both institutions and their students (ranging from construction grants to tuition aid).

5.4.2. Capital equipment

Altering which providers have access to what kinds of capital equipment can have a major impact on provider mix. How government spends money to build and equip its own facilities is the most obvious way to "turn" this control knob. In addition in some countries government is a major source of capital even for private providers. In Germany almost all hospital capital investment comes from the state governments. In a number of middle income countries, such as Thailand, Colombia, and the Philippines, governments have initiated subsidized lending programs to encourage private health facility development. In these situations, various kinds of planning processes – as well as political deals – generally influence the pattern of resource allocation.

Where the private sector is self financing a variety of regulatory and payment tools are still available – in effect using one control knob to change the setting on another. For example, there are states in the U.S. that have largely limited the use of certain technologies to hospitals – thereby insuring that free standing physicians groups cannot offer particular services such as radiation therapy or MRI diagnosis. Such regulation is justified by a concern over clinical quality and hence health status results – as well as by a desire to constrain total costs.

As mentioned previously, deciding to pay for certain services only in some situations and not others can also have a significant effect on macro-structure. For example, the U.S. Medicare program will not pay for long term psychiatric care in purely psychiatric hospitals. This decision was made self-consciously to change the mix of providers that offered this service – and helped contribute to the enormous decline in hospital beds in psychiatric hospitals in the U.S. in the 1970s.

5.4.3. Pharmaceuticals

An important and growing arena for input-control policies is in the area of pharmaceuticals. Policies limiting imports and/or domestic production to certain "essential drugs" are aimed directly at the production function for health services (which as we noted at the beginning of this chapter, is an aspect of micro-organization). Again, where government directly operates the system it can implement an essential drugs program through its own ordering and stocking behavior. To control private sector

pharmaceutical use requires it to use regulation to restrict imports etc. Most recently, the international debate has turned to questions of increasing access to certain drugs, such as those needed to treat HIV/AIDS. Access to these drugs at much lower prices would certainly affect the micro-organization of HIV/AIDS treatment providers…and possibly macro organization as well.

Governments also regulate the right to sell drugs and other items as a device for changing provider mix. In many countries only licensed pharmacies can dispense certain classes of compounds. And the conditions on obtaining a license (e.g. training requirements) in turn restrict the nature of the enterprises offering the service. Indeed some European countries restrict the sale of certain profitable non-prescription items to licensed pharmacies in order to cross-subsidize those providers and hence make possible a denser network of sellers.

In sum, government has a variety of tools to change the distribution of providers, or their production technology, by controlling the availability of inputs. This changes the capacities of workers, or the system of production, and may also effect what providers provide what services. Most of these methods operate relatively slowly – since they often act more on the "flow" of new inputs than on the "stock" of available inputs. And when the stock is long-lived, change takes time. Still efforts at input control may well be a critical part of a portfolio of interventions designed to change the micro and macro structure of the health care system.

6. Concluding Observations

Our discussion of the organization control knob has been relatively long because so many different ideas and actions are subsumed under this heading – everything from decentralization and competition to privatization and contracting. Basic issues of provider mix, public and private sector roles, hospital structure, etc. all come under this rubric.

Throughout our discussion we have tried to keep in mind a basic conceptual approach that we believe will allow readers to both diagnosis causes of poor performance in their own country and to devise appropriate organizational responses. That approach focuses on two related effects: the consequences of system level changes on performance as well as the effect of changes that impact on the individuals in the organizations – on their incentives, attitudes and capabilities – and the consequences these have for performance. It particularly focuses on managers, whose actions in turn shape so much of what occurs. And we urge that the key impact of policy interventions like decentralization or autonomization can best be explained by understanding their impact at the level of individual managers and workers.

The effect on individuals however is not the only link from various settings on the organizational control knob to performance consequences. Provider mix has an impact on scale and skill. Some kinds of facilities and some kinds of health workers are more attractive and comfortable for some patients than others. Levels of investment and

budgets for supplies all play a role in shaping performance. These forces are important in defining what managers and workers can do. They work themselves out in and through the choices available to, and the decisions made by, front line workers and their managers. All this returns us to an early point. Many organizational changes have very uncertain effects unless and until the details are spelled out. One has to match incentives and capacities – authority with accountability. And implementation is critical. Yet it is often the case that the leaders of the Ministry of Health or the Social Insurance fund or the medical schools are themselves *not* managers. They often have a short tenure and little executive experience. They are not especially interested in the psychologically draining process of direct managerial supervision – with its complex mix of teaching, rewarding, correcting and advising. Since they don't know this is their job, how can they create systems to select institutional managers to do the very tasks that they themselves often ignore or devalue? This again suggests some conditional advice about what it takes to implement some of these changes successfully. These cannot be done in a routine or hands-off way. Reformers need to consider carefully if they have the political support and administrative savvy to turn the organization control knob successfully. Changing organizational systems at either the micro or macro level, so as to actually improve performance on our core criteria, is not an enterprise we recommend for the impatient or the faint-hearted. Yet it also may well be essential if other reforms (especially financing and payment) are really to yield their hoped-for gains.

Chapter 10
Regulation

1. Introduction

In one sense, we can view both government and market as means to achieve societal goals. Nations can choose to rely on the free market to finance and deliver health care, including the production of drugs and education of health professionals. States can also choose to have government directly handle these functions. Both alternatives constitute a mean to achieve society's health ends. Which alternative is better largely depends on what ends a society aim for. Considerations in making this policy choice – providing health care through government organization or a free market are discussed in Chapter 8 (macro-organizations.)

The current enthusiasm among Western-trained economists for using markets in the health sector reflects their implicit commitment to what we have identified as a "subjective utilitarian" philosophical perspective. That perspective takes as its goal increasing the sum total of everyone's utility as determined by the individuals themselves. Subjective utilitarians make a strong assumption that as income is already distributed in some "satisfactory" manner, then competitive markets are the best vehicles for individuals to maximize their own satisfaction (i.e. utility).

Whenever a nation decides to rely on the free market to perform any of the major functions in financing and production of health care, it has to regulate one way or another. One basic reason for this is that market exchanges require rules governing property rights and for the honest and open exchange of services, labor and goods (North, D.C. 1990, Fuller, L. 1969); otherwise, the most manipulative, powerful and/or strongest persons will take all (Oakeshott, M. 1975). In civilized societies, then, *the state imposes constraints on organizations and individuals (Heriot, G.L. 1997). This is called "regulation".*

Regulation is a major control knob for health systems. It can influence the allocation of resources when it specifies what health services must be included in the insurance benefit packages, thereby affecting health status and public satisfaction. It can also set rules that enable the insurance market to operate more competitively, thereby improving efficiency, which in turn affects risk protection and health status.

Regulation is also intimately connected to the other control knobs: for these to be effective, they must be accompanied by appropriate regulation. For example, a country can establish a social insurance scheme, but the it is likely to go bankrupt unless effective regulation compels the eligible population to enroll and pay premiums. By the same token, regulation is much more effective when combined with the appropriate incentives, education, and macro-organizational arrangements.

Because regulation is a powerful control knob, it must be used sensibly and appropriately – inappropriate regulation can make health systems worse rather than better. Policymakers must therefore understand the strengths and limitations of regulation.

This chapter is organized around five questions: What is regulation and why regulate? How do regulations influence the performance of health systems? What are the key considerations for regulatory success? What are the major types of regulation in the health sector? What are the principal regulatory failures and our conditional guidelines? The fourth section – a description of major types of regulations is rather long because the myriad of regulations and they differ greatly between high and lower income countries.

2. What is Regulation and Why Regulate?

2.1 What?

A state may use legal rules, economic incentives, and persuasion to influence individual and organizational behavior. Regulation, in a narrow and clear sense, refers to the government's use of its *power to impose constraints* on organizations and individuals. Under this definition, only legal rules, and not incentives or persuasion (see Chapter 10), are considered to be regulation. Regulation, then, includes the full range of legal instruments, (e.g., laws, decrees, orders, codes, administrative rules, guidelines), whether issued by the government or by non-governmental bodies, (e.g., self-regulatory organizations) to which the government has delegated regulatory power (OECD, 1996).

Our definition of regulation also excludes rules and requirements imposed by organizations that purchase insurance or health services on behalf of a group of people. For example, in order to be prudent purchasers, social and private insurance plans often set rules and restrictions on providers. These are not government-imposed legal rules, but rather contract terms negotiated and agreed upon between buyer and seller. This holds true even when the government acts as the purchaser, contracting insurers or providers for services.

2.2. Why?

What are the rationales for regulation? When is it justified for the government to use its power to constrain the freedom of individuals or organizations? The justifications for regulation depend on *ethical beliefs* (presented in Chapter 3) derived from and determined by a nation's social values (Shkar, J. 1997).

In general, four reasons are most frequent d as justification for the use of state power. First, any society relying on market exchange mechanisms must ensure that exchanges and transactions are done honestly and openly (North, D.C. 1990, Williamson, O. 1987). Property rights have to be defined and protected; dishonesty, deceit, and fraud, prohibited. But even when markets work well from an economist's perspective, markets can't produce all the outcomes that matter to a society. The second reason, therefore, is that while markets can improve efficiency and produce goods demanded by consumers, *markets can't deal with unequal distribution of income and varying health needs*. Third, health-care markets suffer from major market failures, regulatory interventions are necessary to remedy these failures in order to maximize consumers' utility. Lastly, as followers of communitarianism and objective utilitarianism place greater values on selected outcomes, they would regulate markets, because they sometimes do not like the results that markets produce, such as abortion service.

When any society relying on market exchange mechanisms, it must ensure that property rights are defined and protected, otherwise honest exchanges can't take place. In short, the rules for honest game in transactions and exchanges must be set via regulation; these rules define the legal obligations of the various players in the transactions, delineating their power, responsibility and accountability. In the health sector governments regulate the obligation and responsibility of both buyers (patients and consumers) and sellers (doctors, pharmacies, hospitals, and insurers) to ensure that agreed-upon exchanges and transactions are honest, transparent and executed completely. Usually, the particular standards of honesty and transparency used derive largely from the ethical beliefs of that particular society.

Free markets yield efficient results, but exacerbate any inequity that already exists in the society. Health care can't be produced and distributed equitably unless households have equal purchasing power for health care and have similar health conditions – a condition that does not exist in reality, of course. As a result, poor citizens get fewer services from hospitals organized on market principles, precisely because they are poor and cannot afford to pay. By the same token, few doctors locate in depressed rural areas because of the limited economic opportunities these areas offer. Insurance markets too can produce objectionable results. For example, without regulation, such markets often will not cover chronically ill people because they can not afford the premiums. All of these situations can and have produced regulatory responses in countries rooted in egalitarian traditions, where regulations are used to promote more equal access to health care and distribution of health status.

Health-sector markets often do not possess the prerequisite conditions for reasonably effective competition. Economists call them "market failures" because they inhibit utility maximization; hence, the justification for regulatory intervention. For example, if consumers are unable to judge the quality of medical care, and might therefore be victimized by unqualified practitioners, government can impose training or licensing laws to protect them. But for economists, the goal remains that of utility maximization. When patients are unable to make informed purchases of drugs in situations where drugs may have missing or impure, or harmful ingredients, other than

the advertised ingredients. Yet, the patient can't tell the difference. Government regulates the purity of drugs to protect the health of the general population. Enforcing this regulation requires the government to set standards for pharmaceutical production processes and to test randomly selected samples of drugs for chemical purity. Then, police power, a legal process, and court system are needed to enforce the regulation, adjudicate disputes, and impose penalties.

Regulations must also be used to remedy serious market failures such as monopolies. Regulation in the health sector is very extensive in all high-income nations, though its extent, substance, and organization vary enormously.

Besides correcting market failures, government also uses regulatory interventions to assure public goods, such as clean water and sanitation, are made available. Regulations are also frequently introduced to reduce negative externalities caused by individual actions, such as rules for road safety and firearms.

Communitarians and objective utilitarians can also find reasons to object to market results. For instance, consumers may not voluntarily purchase cost-effective preventive services, so that the market does not maximize health status, or they may seek to purchase services (e.g. abortion) that communitarians object to on moral grounds. The influence of ethical beliefs on regulations can be strong. For example, a fervent objective utilitarian would favor regulating specific individual behaviors, (e.g., the quarantining of recalcitrant HIV infected individuals who engage in unsafe sex), as long as such regulation would improve the aggregate health status of the population: the ends (better health) would justify the means (regulation to change behavior).

Libertarians, however, would oppose any efforts to limit individual autonomy, even if this respect for individuals would result in unhealthy behaviors leading to less favorable aggregate health status. A communitarian, on the other hand, would applaud coercive techniques in order to alter behaviors that violate cultural expectations and community values. Thus our analysis of regulation does *not* presume either that the market is always best or that the economist's view of society's goals is the only way to look at health-sector reform. Instead, we proceed on the assumption of the legitimacy, and relevance, of the diverse ethical perspectives that in fact inform the reform process in various nations around the world.

In sum, the ethical justifications for these four categories of regulation are summarized in Table 10.1.

Table 10.1.: Categories of Regulation and Their Related Ethical Perspective

Category of Regulation	Related Ethical Perspectives
Establish basic conditions for honest market exchange	Subjective utilitarianism, objective utilitarianism
Perfect what markets cannot do (such as ensure equal access to basic health care)	Egalitarian liberals, communitarianism
Correct market failures and provide public and merit goods	Subjective utilitarianism
Correct unacceptable market results (such as use of tobacco)	Objective utilitarianism, communitarianism

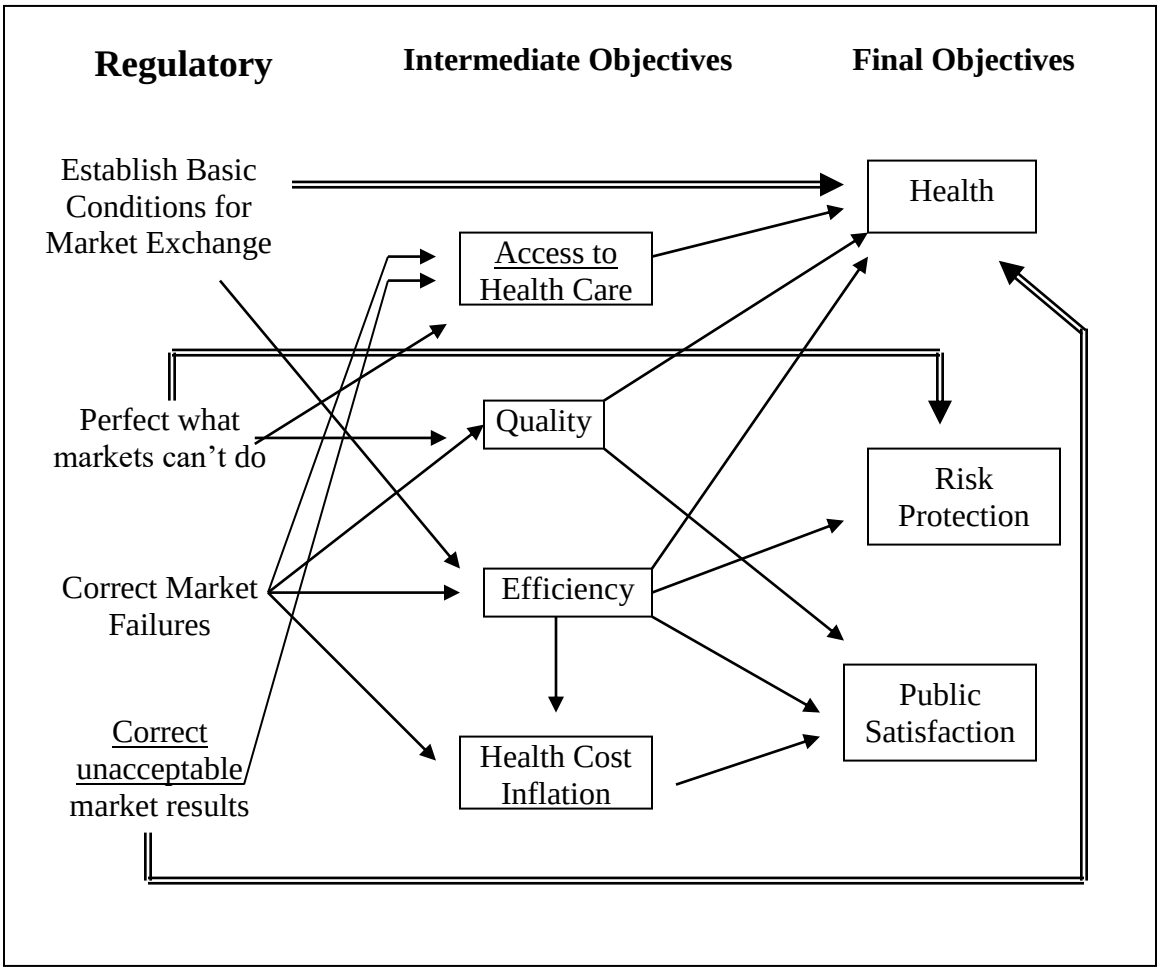
3. Regulation and Health System Objectives

How are categories of regulation connected to the intermediate and final objectives of health system as presented in Chapter 4? We illustrate their relationships in Figure 10.1.

Only certain regulations can directly affect the core outcomes of the health system; these have major impacts on a population's health or provide financial risk protection. For example, ensuring the safety and purity of water, food, and drugs because of consumers' inability to judge the quality of these goods. Another example is the regulation of compulsory enrollment in social insurance plans to ensure every eligible person has financial risk protection.

By contrast, most regulation impacts intermediate rather than core objectives. For example, public safety regulations aim at improving the quality of health services, which in turn improves health outcomes and/or public satisfaction with services received (see Figure 10.1.). Regulating market failures can improve efficiency in the provision of health services and insurance, thereby lowering health costs and making health services more affordable. Regulating monopolistic pricing can lower the prices of health services, giving patients greater access to health services, in turn affecting health outcomes and patients' satisfaction with the health system.

Figure 10.1. Relationship between types of regulation and health system objectives



4. Key Considerations for Effective Regulation

In this section, we discuss several key considerations for the successful adoption and implementation of regulations. While having a strong reason is a necessary condition for establishing regulation, it is not a sufficient unto itself. Effective regulation also requires *good design and wording* and governmental *ability to enforce* the regulation. Lastly, we briefly discuss the importance of political feasibility. Politics determines regulatory policy; regardless of how technically well designed a regulation is, it must be politically acceptable. The relative importance of the four considerations for regulatory success varies among nations, depending on their culture, political institutions and process, and public governance.

Before proceeding, we must point out the differences in conditions between high-income and low- and middle-income countries. Most low- and middle-income countries lack: (a) adequate expertise to promulgate regulations; (b) administrative capacity, police,

and courts to enforce regulations; and (c) information to monitor compliance. As a result, high- and low-income countries regulate health services and insurance very differently. There are fewer regulations in lower income countries, and those that exist focus mostly on market failures that impair public safety. Enforcement of regulation is uneven, and can be hampered by corruption.

Many high-income countries have the capacity to regulate medical inputs by ensuring the appropriate proportion of specialists to generalists, a reasonable level of new investment for health facilities, and rational diffusion of medical technology. They also establish extensive regulations to assure quality and safety of health services. Moreover, when high-income countries rely mainly on the private sector to provide health services, they can usually control monopolistic prices set by private providers and pharmaceutical companies, while insurers are regulated to assure the solvency of insurance firms and to reduce risk selection.

By contrast, most low- and middle-income countries do little to regulate of the private health care or private insurance markets. Instead, governments turned to direct public provision of health services, managed by bureaucratic rules. This practice may have grown out of the recognition that the government lacks the capability to regulate the quality of private sector health services and the monopolistic prices charged for them. However, private sector providers often remain the major source of primary care services because of the low quality of public sector services. By adopting a laissez-faire policy toward private sector providers, government leaves many patients at the mercy of unregulated indigenous doctors, private pharmacies, clinics and hospitals.

High- and low-income countries also regulate to perfect what markets can't do, but to accomplish different goals. For example, most high-income countries have an adequate supply of physicians; regulations in these countries are commonly adopted to limit the supply of physicians because an excess supply would be likely to increase supplier-induced demand, resulting in a higher rate of health-cost inflation. In contrast, most low-income countries have a shortage of qualified physicians, and their regulations focus on influencing physicians to serve in physician-scarce areas.

Around the globe, numerous regulatory interventions have been adopted. Experience seems to show the successful regulatory efforts depend on four key considerations: competence in designing regulations, enforcement, organization, and political feasibility.

4.1. Competence in Designing Regulations

Institutional competence in design and legislation is a key element of successful regulation. Both technical competence and sound judgement are needed. The substance of a regulation has to be carefully crafted and worded, the penalty has to be commensurate with the seriousness of harm or damage done by violators, and sanctions or incentives have to be skillfully composed to promote compliance; otherwise, the regulation is unlikely to be effective. These tasks require experts from many fields. For example, deciding on appropriate regulations on pure food requires experts from the legal, public health, medical, biology, chemistry, economics and public management fields. Establishing effective regulation to ameliorate market failures requires extensive knowledge about the marketplace and how the regulated may react to the regulation. This requires highly competent experts from several fields, e.g., law, economics, organizational theory, and public management. Improperly designed regulation can even exacerbate market failures.

In designing regulation, cultural norms and social beliefs must also be taken into account. People's reactions to regulation vary depending on whether or not they agree with the norm expressed by the regulation. When regulation restricting people's behavior is compatible with their social beliefs, they are more likely to comply; otherwise, the regulated may try their best to evade the regulation. Because of this, regulation should often be accompanied by persuasive measures (see Chapter 10) that bring social beliefs into alignment with regulatory intent.

4.2. Enforcement

The actual outcome of regulation critically depends on the enforcement process; specifically, on that process' ability to detect and penalize violators, and to influence the rate of compliance through deterrence.

Because regulatory agencies typically have only *limited enforcement resources,* they can only detect, apprehend, and prosecute a limited number of violators. Only if voluntary compliance is relatively widespread, and violations relatively infrequent, can these limited resources be sufficient to apprehend enough violators so that the probability of detection and prosecution will be high enough to deter potential violations.

To detect violations, *regulators need data*, which creates still other complexities. Often, the data that is easily (and inexpensively) available is not the data that regulators would most like to have. For example, a country that wants to regulate quality can easily test new medical graduates on their knowledge of anatomy. It is much harder to assess how practicing physicians use their knowledge in the daily treatment of patients. Thus, regulators often face a tradeoff between data's relevance and cost of acquisition.

Once violators are detected, sanctions must be imposed. However, if the process for imposing penalties requires long and complex court proceedings, enforcers may not have the resources to prosecute everyone. On the other hand, if enforcers can directly

impose fines without lengthy court proceedings, they may abuse their power. Once police, prosecutors, or judges are involved in the enforcement process, they will inevitably find room to exercise their own discretion about how it is carried out. If rules seem unfair or inappropriate to them, they will find ways to avoid enforcing them.

Together, these phenomena imply that rules and sanctions that are seen as socially acceptable in a given culture are much easier to enforce. High levels of voluntary compliance make the job of rule-enforcers that much easier, and the deterrence effect of enforcement that much more potent.

4.3. Organization

Enforcement is often a complex task. The question of who might most cost-effectively enforce the rules is not easily answered. Responsibility and authority can be given to a government agency, a parastatal organization[24], or a private organization. Which option is most effective and efficient depends largely on the relative managerial capability of these organizations and on the given nation's political institutional structure and process.

Self-regulation has also been used in the health sector to assure quality that is difficult to measure or ascertain (De Geynt, W. 1995). Indeed, health-care regulators often have to rely on the regulated to report on their own behavior, just as tax collectors rely on individuals and businesses to file tax returns. For example, what is the surgical complication rate or infection rate in a hospital? In such cases it is in the regulator's interest to ask for reports based on data already being collected by the regulated for their own purposes. Using such data lowers the cost of reporting for the regulated, making it more likely they will do so. And it potentially gives regulators a place to check on the accuracy of the reports they receive. To encourage accurate self-reporting, regulators have to refrain from penalizing the regulated too heavily for small violations of the rule: if they try to be draconian, they will only encourage inaccurate reporting, resulting in unreliable data.

[24] Parastatal organizations have often been used because of their relative insulation from direct political influence.

Table 10.2. is a matrix of the organizational alternatives that can administer regulations, and some examples of these regulations.

Table 10. 2. Alternative Organization of Regulatory Agencies in Relation to the Organizations to be Regulated

Regulated → **Regulator ↓**	State	Parastatel	Private
State	X	Regulation of social insurance plans such as sickness funds	Government regulation of private companies
Parastatel	X	X	Sickness funds regulate private providers; joint commission on hospital accreditation regulate hospitals
Private	X	X	Self-regulation by medical association

4.4. Political Feasibility

Whether they are legislated or decreed by executive order, regulations are enacted through political processes. All nations have organized interest groups, whether formal and/or informal. These exist whether the government is governed by democratic process or by a dictator, and interest groups can have very powerful roles in influencing regulatory decisions. *Hence ,in designing regulations, political skills are of great importance.* Chapter 5 describes the political skills that are needed, such as understanding how a country's institutional and political structure affect how regulation can be established and enforced, who might support and who might oppose the regulation, what political alliances can be formed, what compromises have to be made, what political strategies might work, and how regulations can be designed to be politically acceptable.

5. Major Categories of Regulation

For clarity of presentation, we divide regulations into two types: health activities and insurance. For ease of understanding, we group regulations into four categories according to their purpose: establish basic conditions for market exchange; perfect what the markets can't do; correct market failures and provide public and merit goods; and correct unacceptable market results. Table 10.3 summarizes the common regulations in

each category; Table 10.4 summarizes the major categories of regulation for insurance. Examples of regulatory instruments are given for each category.

Table 10.3.: A Comparison of Regulations of High- and Low-Income Countries in Public Health and Health Services

Category of Regulation	High-Income Nations	Low- and Middle-Income Nations
1. Establish Basic Conditions for Market Exchange	a Define and protect property rights and patents b Govern solvency and bankruptcy of health service institutions c. Protect patients' rights	a. Similar, but less enforced b. Similar c. Very little formal regulation
2. Perfect What Markets Can't Do (such as assuring equitable access)	a. Assign new medical graduates to serve in underserved areas b. Assure patients' rights to emergency services	a. Similar, frequently done b. Public hospitals provide emergency services
3.CorrectMarket Failures and Provide Public Goods A. Provide public and merit goods	Direct government provision of free or highly subsidized programs, such as health education and immunization	Same
B. Provide consumer information for making informed choice	a. Labeling b. Regulate truth in advertising c. Restrict physicians from advertising	a. Much less regulated b. Much less regulated c. Little restriction
C. Buyers Unable to Judge Quality: (i) Regulate inputs	a. Establish standards for food hygiene and purity of drugs. b. Licensing of physicians, nurses, pharmacists. c. Accreditation of laboratories, hospitals.	a. Similar, but less enforced b. Similar, but may not require periodical relicensing c. Same, but may not require periodical re-accreditation
(ii) Regulate process	a. Establish practice guidelines b. Establish clinical audits	a. None b. Few countries have it; public hospitals rely on admin. rules.
(iii) Regulate outputs	a. Establish standard quality report cards b Establish liabilities for medical negligence malpractice	a. None b. Most countries have no specific medical malpractice laws, but covered under other common laws (India an exception)
D. Providers' Ability to Induce Demand	a. Limit training slots and "billing" numbers.	a. Most countries have shortage of physicians, so policy is opposite of

(i) Regulate manpower	b. Restrict foreign medical school graduates. c. Disclose conflicts of interest	those in developed countries. b. Encourage foreign graduates. c. Few countries have it.
(ii) Regulate Capital Investment	a. Limit new technology and new facility construction. b. Restrict imports of Equipment.	a. Most countries encourage private investment. b. Some encourage imports.
E. Monopoly (i) Restrict monopoly	a. Enact anti-trust laws and restrict predatory conduct	a. Few countries have it.
(ii) Regulate Monopolistic Price	a. Establish price schedule. b. Establish reference prices for drugs.	a. Establish user fees for public facilities, but not private. b. Very few
4. Correct Unacceptable Market Results	a. Prohibit the sale of tobacco to minors. b. Prohibit assisted suicide	a. Some, but not strictly enforced. b. Very few

5.1. Public Health and Health Services

Here we will describe in more detail the common regulations found in many nations for each category and show the differences between high- and lower-income nations:

5.1.1. Establish Basic Conditions For Market Exchange

Laws are enacted to protect private property rights such as patent laws. In low- and middle-income countries, however, the enforcement of these regulations has been difficult and very uneven.

5.1.2. Perfect What Markets Cannot Do—Enhance Equitable Distribution

A free market bases the purchase of health care on what people can pay and willingness to pay. While, economists have found that the market can produce efficient results, this outcome is not equitable unless the initial distributions of income and health needs are evenly distributed among the people. In reality, of course they are not. To enhance equal access to health services, regulations are used to allocate financial resources to regions where there is greater need. South Africa, Canada, and the U.K. have allocation formulas for public funds that favor regions where the population's health status is relatively worse and people are relatively poor. To equalize human resource distributions, low- and middle-income countries often compel graduates of medical and nursing schools to serve for two to three years in underserved areas.

5.1.3. Correct Market Failures and Provide Public and Merit Goods

In this category, government intervenes in the market for two economic reasons. First, certain beneficial health care is what's known as "public and merit goods", and private firms would not produce them. As a result, some governments compel private firms to incorporate these goods in their products; an example of such is health-education programs. Second, the prerequisite condition for competitive markets is absent unless the government takes action. Health-services markets have several major market failures arising from patients' ignorance of medical affairs and providers' monopolistic power.

Regulatory interventions are designed to remedy three major kinds of market failure: buyers do not have adequate knowledge to make informed choices, buyers are unable to judge quality of health goods and services, and strong monopolistic power of providers. We will here go into more detail about the regulatory interventions for public and merit goods and for the three kinds of market failures.

5.1.3.1. Provision of Public and Merit Goods

The major purpose of this type of regulation is to provide public and merit goods, and to reduce negative externalities. This type of regulation runs the gamut from water to sanitation, immunization to quarantines, and health education to firearms.

5.1.3.2. Buyers Cannot Make Informed Choices

Buyers are unable to assess the content of food and drug products to know their biological effects. Some producers of pharmaceuticals, herbal medicines, and food products profit by exploiting buyers' incapability of knowing about their products. Regulations are needed for two purposes: to inform the buyer of the ingredients, and to ensure that products actually contain the ingredients claimed by producers. Nearly every country requires producers to disclose the basic ingredients contained in a good on the packaging. High-income countries devote significant resources to ensure that products actually contain the ingredients and their purity, but most low- and middle-income countries are quite lax in enforcement due to lack of resources and inadequate administrative capability.

Nor do most patients have the necessary knowledge required to judge the appropriateness and technical quality of medical services and drugs. Misleading information can be given to patients which leads to harmful results. As a result, most countries regulate advertisements by physicians, pharmaceutical companies, and hospitals to restrict possibly misleading information. Government also regulates what information should be disclosed to patients.

5.1.3.3. Buyers Cannot Adequately Judge Quality

At the same time, patients are ignorant about medical science and can't make informed judgments about what tests and services they should buy and the quality of those services. Governments are particularly concerned that some private for-profit providers may give poor-quality medical services and harm the patients. A range of regulations on inputs, process, and outcomes has been used to this end.

Quality of services has two aspects – technical competence and professional caring. P[atient satisfaction has become a key aspect of quality evaluations], and p*atients' assessment* focus on the "caring" rather than the technical component of health services. Nevertheless, patients can also provide information as to what was done technically, as for instance whether a particular necessary examination was performed before a diagnosis was made, whether there was a follow-up on laboratory tests, etc. In the U.S., patient and

community involvement in assessing both technical and personal quality of health care has increased significantly. By feeding this information back to physicians and other health professionals, management has used patients' assessment successfully to improve the quality of services.

While quality assurance was once left to health professionals, the effectiveness of self-regulation has come into question, and the governments of more advanced economies are taking an increasingly active role in regulating and monitoring the quality of medical services. Whatever method is used to assure quality must, however, be credible and acceptable to the medical profession and linked to incentives. The mechanisms used to assure quality include measures such as accreditation of medical professionals and facilities, clinical audits, use of practice guidelines, monitoring of outcomes such as surgical mortality rates, disciplinary procedures, and managerial practices. These measures work for both private-market and government-run systems.

(i) Regulation of Inputs

(i) a. Drugs

Safety and efficacy concerns are the main impetuses for the regulation of drugs. Governments are committed to consumer protection—because of the potential for toxic or other negative side effects, all countries today have regulations that keep drugs from the market until they are proven to be safe. Regulatory bodies also require drug manufacturers to demonstrate the drug's efficacy (i.e., that it performs as described), and in some countries its cost-effectiveness. Manufacturers must also provide accompanying information about possible adverse reactions and contraindications, as well as potential reactions when combined with other drugs.

The U.S.'s market-driven system has established one of the most rigorous approval structures for drugs, with an extremely high level of safety, but with a correspondingly high cost. The Food and Drug Administration (FDA) requires three phases of tests prior to approving a new drug: a) animal research; b) randomized controlled trials; and c) clinical investigations. Once past this prescribed testing, the drug may be approved for routine use by any physician.

Health services are also heavily regulated because of usual consumer ignorance about the etiology of disease and the sophisticated science that underlies modern medicine. The health of the population depends significantly on the decisions made by health professionals and facilities. It is much more difficult, however, to regulate health services than to regulate the safety of a physical product. In health services, the government can regulate either the input, the process, or the outcome. Regulation of input is easiest to establish and enforce, so it is the most prevalent; it does not assure better health outcomes, however. Regulating the process used to produce health services is much more difficult, and the responsibility to self-regulate is often given to the medical profession. Regulation of outcome may seem ideal at first, because it is the outcome with which a society is most concerned; however, outcome measures do not necessarily reflect

the competence and quality of health services. One reason for this is that the same disease can have different degrees of severity, and patients may respond to treatment differently depending on the degree of severity. In general, the medical profession strongly resists regulation based on outcomes, except in extremely unambiguous circumstances.

Many approaches to regulating health services have been developed and used in high-income countries. These are presented below. Low- and middle-income countries have largely relied on accreditation, only because the implementation and enforcement of other approaches would require sophisticated medical expertise and organizational capacity.

(i)b. Licensing and Accreditation

Licensing and accreditation are common tools used to ensure a basic level of competency for health professionals. Nations often combine high accreditation standards with requirements for regular re-certification. For health facilities, the accreditation process generally focuses on structural measures. Once basic quality levels are met, accreditation does little to encourage further improvements. Organizational accreditation takes two forms: either a) the government takes a direct role in setting standards and uses them as preconditions for continued operation and funding; or b) an industry's self-regulating body defines and monitors the standards of institutions that voluntarily choose to participate in the scheme.

India, Egypt, and Nigeria offer typical examples of the licensing of physicians to practice medicine. A graduate of an accredited medical school applies to the government board for a license to practice by showing satisfactory completion of the required courses and evidence of good character. Once s/he is granted a license to practice, no re-licensing is required: authority and responsibility for ongoing regulation and monitoring rests with the medical councils (Bennett, 1994).

In high-income countries, licensing and accreditation of physicians are two independent processes. A government agency is responsible for licensing, while accreditation is often achieved through a professional peer review system. For example, in Canada accreditation is conducted by an autonomous, independent board, with representatives from the medical profession making up less than 50% of board members. This organizational structure has led to greater emphasis on education and self-development. In the U.K., the main accreditation procedure is professional peer review through clinical audits; the emphasis is on self-improvement rather than the removal of aberrant physicians.

The licensing of private health facilities usually requires their physical buildings to meet certain standards and their practitioners to have particular qualifications. In Kenya, for example, a private clinic must stock essential drugs, keep an accurate record of all drugs, and be housed in a non-residential building in a good state of repair (Bennett and Ngalande-Banda, 1994). The U.K. has separate regulatory structures for the private and public health care systems. The King Edward's Hospital Fund, an independent

foundation in London that aims to improve the quality of management in the National Health System, has developed a partial accreditation system. The U.S. relies on self-regulatory initiatives taken by health care provider organizations and by the medical profession. The Joint Commission on Accreditation of Healthcare Organizations (JCAHO) evaluates, an independent not-for-profit organization, and accredits over 18,000 health care organizations and programs. Federal and state regulations largely rely on JCAHO standards. However, JCAHO is reluctant to deny accreditation except where there is a clear and serious deficiency, because insurance plans only pay for services rendered by accredited health facilities. Any organization that loses its accreditation faces potential financial ruin.

(ii) Regulation of Process

(ii)a. Practice Guidelines

Practice guidelines are used to influence clinical decision-making. Theoretically, guidelines should be based on a systematic review of evidence, and sometimes complemented by the opinions of medical experts. Practice guidelines can be effective in creating a standard of care when physicians both view them as clinically credible and face compatible payment incentives.

There is evidence that dissemination of guidelines alone is not enough to alter practice measurably. For example, the state of Massachusetts decided to reduce hypoxic brain damage from anesthesia; it also issued a regulation stating that anesthesiologists could not be found negligent or liable for malpractice claims if they followed the standards promulgated by the American Society of Anesthesiologists for monitoring patients under general anesthesia. After implementation of these new rules, no episodes of hypoxic brain damage occurred .

(iii) Regulation of Outcomes

Quality assurance is most effective when outcomes are monitored. Studies find that quality assurance measures such as accreditation and practice guidelines have little statistical association with measurable outcomes such as mortality rates. Four factors are recognized as basic ingredients of an outcome management system: a) reliable outcome measures (e.g., complication rates and unplanned readmission rates); b) standards for medical procedures; c) comprehensive and reliable databases; and d) adequate analysis. While a great deal of information is available, systems for processing the data are costly and require sophisticated algorithms that can handle the myriad of data .

Sometimes, publishing outcome results alone can improve quality. For example, the publication of New York's data on each hospital's surgical death rate due to coronary artery by-pass surgery was associated with a 41% decline over four years in operative mortality that was not due to patient selection. There are reports that some hospitals took direct action to reduce mortality rates in response to the publications .

In most nations outcome data has been collected but not disseminated because of strong physician opposition. Scotland is an exception. The Scottish Office published data on emergency re-admissions, mortality after admission for stroke, cervical cancer mortality and childhood incidence of measles. Though the publication was issued with a warning that no direct inference about the quality of care in particular hospitals could or should be drawn, some purchasers reported changing contracts as a result of the data release. In the future Scotland may publish data on conditions for which there are large unexplained variations among areas or hospitals .

(iii) a. Clinical Audit and Peer Review

The U.K. and its former colonies rely on the clinical audit system to assure the technical quality of medical services given. The effectiveness of this has not been made clear to the public. In the U.S., managed care organizations report that successful quality assurance requires a multi-pronged approach. The mechanisms found to be useful include preceptorships, audits and feedback, clinical opinion leaders, academic detailing, reminders, local consensus procedures, and patient-mediated interventions.

(iii) b. Disciplinary Boards

A majority of countries rely on the medical profession to discipline physicians. For example, the British General Medical Council, a statutory professional organization, is responsible for licensing, as well as for receiving and processing complaints against the medical profession. The disciplinary efforts of the Medical Council have been highly uneven, however, and substantial barriers to comprehensive and universal programs to protect patients' interests exist.

Some countries create state medical licensing and disciplinary boards, empowered to process complaints against and discipline doctors. These have been largely ineffective, however, due to the domination of these boards by the medical profession. Disciplinary proceedings for professional misconduct by medical practitioners may impose the ultimate sanction – the revocation of the license to practice. But this penalty, which would deny a physician the ability to earn her livelihood from medical practice, is rarely imposed.

5.1.3.4. Moderate Monopolistic Practices – Control Induced Demand

Empirical studies have documented that, due to their superior medical knowledge, physicians, hospitals and pharmacists have the market power to induce patients' to demand for additional units of their services or for goods that have little benefit for patients. The Providers may induce demand when there is insufficient patient-initiated demand for their services or goods. Increased quantity of services resulting from induced demand raises total health expenditure without commensurate benefits to patients. In order to control increases in health expenditure, many high-income countries try to regulate medical manpower and investment in new medical facilities and equipment.

(i) Regulate Manpower

When a country has an oversupply of physicians, total health expenditure increases more rapidly, because physicians can strongly influence patients' demand for health care. When physicians experience insufficient demand for their services, they can increase the volume of services by advising patients to come back for additional visits, by ordering more tests, or by performing surgeries on patients who may not need them. The number of physicians has been positively associated with quantity of services used, which in turn leads to increases in volume of health care and expenditure. For these reasons, markets cannot be relied upon to adjust supply to be commensurate with demand, and most developed economies find it necessary to regulate the supply of physicians. There are three points at which this regulation can be applied: entrance to medical school, entrance to residency programs, and licensure.

Most high-income countries limit supply by controlling the number of students admitted to medical schools. Some nations also use the number of residency slots as a means of obtaining the desired quantity of physicians in primary care and in each medical specialty. For example, the U.K. Department of Health prescribes the total annual intake of candidates to medical schools. Since the 1940s, committees have advised on the number of physicians to be trained, alternating between reductions due to concerns about unemployment in the medical profession and increases to utilize medical school capacity.

National governments can directly control the supply of physicians – both total number and specialty mix – by regulating the number of slots in medical school and/or postgraduate residency programs. There are also a variety of indirect methods; for instance, the size of budgets given to medical schools, and the amount of scholarships awarded to medical students. In a direct approach, Canada conducted a nationwide study of medical manpower needs in 1975, based on the *need ratios* of various specialists to the population. The resulting report was developed with considerable medical school input. Based on the report's findings, medical schools were asked to adjust the number of specialty slots offered and to try to influence medical students' choices regarding specialty and geographic location. Singapore limits supply by restricting specialists to 40% of new physicians and training the remaining physicians to be primary care gatekeepers. Singapore also restricts licensing of new foreign medical graduates to a very low number each year.

High-income nations also use regulations to redistribute human resources. For example, when Canada predicted that it would have a surplus of physicians, federal authorities placed restrictions on the immigration of foreign doctors, requiring those admitted to serve physician-scarce areas.

Limiting supply, however, can reduce competition. In the U.S., specialty boards have been able to push up prices by constraining the number of specialists who become board certified to practice in a particular specialty such as anesthesiology.

While high- and middle-income countries are concerned about oversupply of physicians and their maldistribution by specialty, low-income countries are mainly concerned about the inadequate supply of qualified doctors. These nations try to increase their number of medical school training slots, and their regulatory policy focuses mainly on redistributing qualified physicians to rural and remote areas. For example, Egypt and Ecuador assign and require new medical school graduates to serve two years in physician-scarce areas.

(ii) Restrict Capital Investment

In the face of the growing tension between affordability and availability of services, prioritizing investment for health interventions is increasingly important to ensure that scarce resources are used in ways that achieve the best health outcomes. As capital investment and technology acquisition increase, more resources are dedicated to specialized care, and fewer to prevention and primary care. If we follow the objective utilitarian tradition and are concerned about the cost-effectiveness of health services, this calls for regulation of capital investments, particularly in new medical technology, to ensure they meet the nation's standard for cost-effectiveness.

Regulation of capital investment requires technology assessment. The process for selecting a technology for evaluation is a central issue in technology assessment, and the majority of Health Technology Assessment (HTA) organizations use several criteria to establish priorities as to which new technologies will be evaluated. The criteria include *potential cost of the technology, prevalence of the medical condition, assessment feasibility, and perceived interest by the public.* However, prioritization efforts can be distorted because new and expensive technology often has strong advocates—those who are in a position to benefit financially from their use.

Technology acquisition is closely linked with manpower considerations, and physicians are a major factor in the demand for and use of technology. Many U.S. hospitals require physicians to have specific training before being allowed to use high-tech devices (e.g., lasers, endoscopy). Medical and professional societies are also taking an active role in assessing and ensuring the appropriateness of technology use.

Under national health service systems, decisions on the cost-effectiveness of alternative spending decisions for new capital facilities, technology acquisition and approval of new drugs can be linked with budgetary and planning processes to ensure that overall health care system objectives are met. Market-driven systems lack such a planning process and must therefore depend upon competitive market forces and governmental regulatory bodies to determine how investment decisions will be made.

5.1.3.5. Monopoly

(i) Restrict Monopoly

Both private and public monopolies have been used to exploit patients—the former for profit and the latter for gains to the public health service practitioners.

Monopolistic hospitals, specialty clinics, and pathology and imaging laboratories can compromise quality of services and charge monopolistic prices. High-income countries establish and enforce anti-trust regulations to curb private monopolies. Internal markets have emerged in the past decade as an approach to curbing public monopolies.

In low- and middle-income countries, both public and private monopolies are widespread, since health facilities and specialists are few. Regulation is ineffective at curbing the monopolistic practices of private for-profit institutions. To curb the monopolistic power of public hospitals and clinics, international experience indicates that non-regulatory approaches (e.g., separating financing from provision and having the "money follow the patient") hold some promise of improving the efficiency and quality of public institutions (see Chapter 8 on organizations).

(ii) Regulation of Monopolistic Prices

Even though physicians' power to charge monopolistic prices has been extensively documented in many countries (Kessel, 1974, Bennett, 1994), all countries find this impossible to regulate directly. Medical practitioners provide thousands of diverse services; regulating fees is therefore extremely complicated, requiring detailed information on each transaction. Bennett et al found that no African government regulates fees charged by private practitioners (Bennett, Ngalande-Banda, 1994.) Furthermore, while social and private insurance plans, acting as purchasers, can set the fees that the plan will pay, they have not been effective in prohibiting practitioners from collecting additional charges from patients. Most insurance plans try to negotiate a fixed fee schedule with umbrella organizations representing practitioners, then leave the organization to these umbrella organizations. The latter rarely exist, however, in low- and middle-income countries.

For most countries, the primary concern is *not* the high prices that private practitioners can charge, but the fact that their higher income creates instability for public sector providers. Practitioners employed in public sector facilities respond in several ways to higher earning potential in the private sector: some resign and leave for full-time private practice, some establish a dual practice and refer wealthier patients from the public clinics to their private clinics, and some begin charging under-the-table fees. All of these responses weaken or corrupt public sector providers. Meanwhile, low- and middle-income countries can hardly afford to pay wage rates to public sector practitioners that can compete with private sector earnings obtained from high charges. As a result, many low- and middle-income countries restrict private practice to physician-scarce areas (Mutungi, 1992).

Besides safety, drugs are also regulated for their efficacy and cost-effectiveness. In both the U.K. and Australia, new applications for a drug to be reimbursed require evidence of safety, efficacy, and cost-effectiveness. Each of these criteria must be based upon findings from objective studies.

5.1.4. Correct Unacceptable Market Results

Results produced by competitive markets are not necessarily acceptable to the objective utilitarian and communitarian. Objective utilitarians who believe maximizing health status is most important, would regulate the consumption of tobacco and other goods that may harm people's health, mandate the use of seatbelts and helmets to reduce injuries. A Communitarians may find blood transfusions, sale of organs for transplantation, abortion, and assisted suicides violate community values. They would regulate these.

5.2. Insurance Organizations

In order for market exchange to be fair, consumers must understand what they are buying. Private insurance is a complex futures contract whereby the insured pays a premium in advance for specific compensation if certain defined events happen; the insurance policy is a legal contract that spells out all conditions, obligations, contingencies, and exclusions. Most of the contract is difficult to understand because it is written in legal and technical insurance language, and consumers can be easily exploited. Most insurance regulations aim to prevent such exploitation and fraud.

Health insurance can be provided by three types of organizations: government agency, parastatel organizations, or private companies. We briefly present and discuss here regulations for the latter two categories. There is little involvement by insurance organizations in the public safety arena. Rather, all of the regulations pertain to: 1) rules of honest exchanges; 2) perfect what market can't do, such as promoting equity; and 3) correction of market failure. Regulations differ depending on whether the insurance is compulsory or voluntary. Under social insurance, where households are compelled to purchase the insurance, the state is more concerned about equity issues, unfair pricing, and deceitful business practices. Regulations for voluntary insurance are usually more stringent. However, in most developing countries, private voluntary insurance companies face few regulations besides the requirement to obtain the initial license to operate.

Table 10.4. summarizes the various types of insurance regulations. Note that insurance is seldom involved with public safety; thus, that category is not shown in the table.

Table 10.4.: Regulation of Insurance

Category of Regulations	High-Income Nations	Low- and Middle-Income Nations
1. Establish Basic Conditions for Market Exchange (A) Solvency of Insurance plans	a) Establish adequate minimum capital and surplus standards. b) Limit investment options. c) Establish financial reporting requirements. d) Establish standard for long-term actuarial soundness for both private and social insurance.	Modest regulation of private health insurance with weak enforcement. Huge profits usually made by companies that are able to obtain a license to sell.
(B) Sales and Marketing practices	a) Advertising. b) Disclosure of commission rates, limit maximum sales and marketing expenses. c) Content and form of insurance policy.	Some regulations but weak enforcement.
2. Perfect What Market Can't Do—Equitable Distribution (A) Risk pooling	a) Require insurance to set premiums on a communitywide basis. b) Compel eligible households to enroll in social insurance plans.	Similar laws for social insurance, but weak enforcement.
(B) Equity in financing and benefits	c) Premium based on a percentage of wages in social insurance.	Similar
3. Correct Market Failures (A) Risk selection	a) Require open enrollment, prohibit medical underwriting. b) Establish risk-adjusted premiums. c) Reinsure high-risk individuals by transferring funds retrospectively from insurers with lower average risks to those with higher risks. d) Require insurance to set premiums on a community-wide basis.	Social insurance is usually regulated, but not private insurance.
(B) Adverse selection	a) Disclosure by enrollee of medical history and condition.	Very few regulations.
(C) Monopolistic pricing	a) Require minimum loss ratio, i.e. a minimum percentage of premiums must be paid out for health service benefits.	Very few countries regulate.
4. Correct Unacceptable Market Results (A) Free-rider	a) Compel all eligible population to enroll in social insurance.	Same, but less effective enforcement.
(B) Cost-effectiveness	a) Regulate benefit package of compulsory insurance	Similar

5.2.1. Establish Basic Conditions for Exchange

5.2.1.1. Solvency *of Insurers*

Solvency problems are less frequent in social insurance systems than in private insurance. For social insurance, inadequate funds are typically addressed through government subsidies, increases in contribution rates, or reductions in benefits. For private insurance, governments regulate to prevent insolvency.

The U.S. is the only high-income country that relies heavily on voluntary purchase of private insurance to fund health care. It therefore has the most experience in regulatory requirements for private insurance. Most insolvency of private insurance firms is due to inadequate pricing, poor investment of insurance reserve funds, and/or fraud. After experiencing a number of bankruptcies where insurers could not meet their contractual obligations, the U.S. pressured states to establish more stringent regulations (states are responsible for regulating private insurance). The states commonly require insurers to meet certain minimum capital and surplus standards, limiting their investment options and establishing financial reporting standards.

Governments also establish minimum standards for the qualifications of those who perform key technical tasks, e.g., the actuaries who determine the premium rates and reserves (Wasem, 1998). The primary tool to monitor the solvency of a firm is review of its financial statements. Solvency regulation encompasses a number of aspects of insurers' operations, including capitalization, pricing, investments, reinsurance, reserves, asset-liability matching, and transactions with affiliates and management .

5.2.1.2. Sales and Marketing Practices

Most private insurance is sold through agents who are motivated through receiving a commission for each sale. Because of the complexity of insurance contracts, private insurance plans and their agents can use deceptive sales practices to sell their products. High-income countries regulate potential deception by regulating the form and contents of insurance contracts. Insurance plans must file their policy forms for approval by regulatory authorities (Wasem, 1995). Sales literature and advertising are also regulated. Furthermore, to prevent private insurance plans from paying excess commission to their sales agents and staff to promote sales, high-income countries establish loss ratio requirements that restrict the percent of premiums that can be used for administrative and sales expenses and profits.

5.2.2. What Perfect Markets Can't Do

5.2.2.1. Pooling Risk Widely

In the voluntary purchase of insurance, the market equilibrium price would be an actuarially fair premium that reflects the expected payout for that individual. This means that the healthy and young would pay small premiums and the less healthy and older

people would pay much higher premiums. To promote equity, many countries require high and low risks to be pooled together for an entire community: instead of having different premium rates for different groups of individuals, premium rates are based on the risk characteristics of the entire community (i.e., community rating). Reinsurance is also used as a mechanism to pool risks across a larger population. Germany requires the premium of all sickness funds to be based on community rating; in addition, there is a reinsurance arrangement to even out the risks across sickness funds . In the U.S., many states regulate Blue Cross, a non-profit health insurance plan, to base its premiums on community rating.

5.2.2.2. Pay Health Premium According to Ability to Pay

Even when a nation uses community rating to pool risk, a flat premium charge would require a low-wage worker to pay a larger portion of earnings for health insurance than a high-wage worker. For equity reasons, many high-income nations regulate by mandating social insurance, and the premium is based on a uniform percentage of wage instead of flat amounts. Columbia has gone one step further. Besides a wage-based premium rate, it adds a surcharge to the rate for workers' social insurance contribution and uses the revenue from the surcharge to subsidize the poor (Londono, 1994).

5.2.3. Correct Market Failures

5.2.3.1. Regulate Risk Selection

Either because of profit motive or the desire to be solvent, all types of insurance plans are likely to select healthy people to insure and reject the less healthy. Governments must regulate this. Under social insurance where consumers can choose to enroll in different insurance plans, governments usually require plans to guarantee the issue of insurance policies to anyone who chooses to enroll during the open enrollment period.

Government regulations reduce risk selection among private insurers with measures such as prohibiting medical underwriting, restricting exclusion of pre-existing conditions, and requiring open enrollment periods. In the U.S., risk selection is a major problem, and in spite of stringent regulations, it has not been reduced to a reasonable level because insurers have numerous subtle ways to discourage the less healthy people to leave and enroll with another insurance plan. New York State had to create a reinsurance mechanism by transferring funds retrospectively from insurers with low risks to those firms with higher risks. To address the problem of risk selection, Australia requires private insurers to set their premium on a community wide basis.

5.2.3.2. Regulate Adverse Selection

When consumers have more complete information about their own medical history and propensity to use medical services they use this superior knowledge to calculate the cost and benefit of an insurance plan offered to them. This is called

"adverse selection". An individual's adverse selection would not allow insurers to pool the healthy persons with the less healthy ones; the premium for the less healthy could then become unaffordable for most. Compulsory insurance has been used to deter adverse selection and to pool the risk between the old and young, the healthy and the less healthy.

5.2.4. Correct Unacceptable Market Results

5.2.4.1. Free-riders

Many nations provide subsidized public health services for those who are uninsured and can't afford to pay for services. These nations may also have compulsory insurance for workers in the formal sector. This arrangement encourages people who are eligible for insurance coverage, but who have to pay a premium, to refuse to sign-up for insurance. To correct "free rider" problems, many nations require that all eligible persons must enroll in an insurance program, and the risks are pooled across a large population to minimize adverse selection. This is one of the major advantages of social insurance.

5.2.4.2. Cost-Effectiveness

In spite of public education efforts, the purchaser of insurance may not have adequate understanding of the benefits of certain preventive services. As a result, the purchaser may not buy insurance benefit packages that contain preventive services. Governments therefore often use regulation to specify what minimum health services must be included in the benefit package.

6. Conditional Guidance

Before we provide any conditional guidance as when to consider the use of regulation to improve health system outcomes, we first want to caution about the extended experience of regulatory failures around the world: this is a major type of government failure that can sometimes make the outcomes worse than if the government had never intervened.

6.1. Major Regulatory Failures

There can be many failures in establishing and executing regulations. Perhaps the foremost form of regulatory failure is corruption. The regulated have a strong self-interest to influence the regulatory agency to adopt practices that will benefit the regulated, including the design and wording of administrative rules and who receives a license to conduct business. Many forms of influence are used, and how the officials of a regulatory agency respond to pressure, persuasion, and bribery depend on the political structure and "professionalism" of the bureaucrats. Corruption is one form of influence, and bribery is the best known form, but it comes in many varieties. Bribes can be offered to regulators to affect their decisions via cash and non-cash gifts, promises of lucrative post-government employment, and favors to the regulators' relatives and friends.

Corruption is, perhaps, the most serious type of regulatory failure in low- and middle-income countries, but it is also extensive in the U.S., though it takes a subtler form. The U.S. has legalized many types of corruption, such as contributions to political campaigns and payments to lobbyists for them to influence, write, and interpret laws/regulations in favor of particular groups or individuals. According to the Corruption Perception Index (based on bribery taking), Canada has been consistently among the least corrupt countries with a score of 9.2, the U.S. has a score of 7.8, and many low-income countries have scores of 3.5 or less with Nigeria being the worst (Transparency International, 2000).

Another key regulatory failure occurs when a regulatory agency is "captured" by the regulated (i.e., those organizations that the agency is supposed to regulate). In theory, the regulatory agency exists to advance and serve the public interest; that may not be the case in real life, however. In the worst-case scenario, a nation's politics may allow the regulated to capture the regulatory agency by having officials appointed to lead the agency who are sympathetic to and will advance the interests of the regulated rather than those of the public (Stigler, G. 1972, Winter 1990, Zellman 1994).

6.2. Guidance On the Use of Regulations

It requires good judgment to use the regulation control knob appropriately to achieve regulatory success. We define success to mean that the regulation largely achieves what the law/administrative rules had originally intended. This implies that the intended purpose of the regulation was clear and unequivocal. We discussed the four major considerations for effective regulation in section IV: technical competence, enforcement, organization and political skill.

It is often, necessary to accompany other control knobs with regulation; otherwise they would be unfeasible or ineffective in achieving their intended purposes. At the same time, regulation by itself has inherent shortcomings, and it often needs other control knobs to supplement it. One reason for this is related to the nature of regulation: most regulations forbid or limit what organizations/individuals want to do – the opposite of incentives. These regulations, by themselves, have limited effectiveness because they

have to rely solely on the coercive power of the state to enforce compliance. Combining persuasion and incentives with regulation would enhance greater voluntary compliance. This consideration seems to have been overlooked by the current push by international organizations for ministries of health to relinquish their financing role and play a greater regulatory enforcement role on a stand-alone basis. Another reason for supplementing regulation with other control knob actions is that laws, decrees and rules have to be to be explicit as to what is allowed or what is not, in black and white terms. But many life events are shades of gray, and regulations are hard to apply in these cases. Combining regulation within centives, therefore, makes rules more flexible and increases voluntary compliance.

There is a great tendency for low- and middle-income nations to simply transplant the regulations of the high-income nations and make them their own, particularly in this age of economic globalization. The establishment of such regulations are easy, fast and inexpensive. Unfortunately, these nations are unlikely to enjoy regulatory success unless they have taken into account the following:

*Country Variations***:** The potential for using regulation to actually change behavior are obviously greater in those countries where popular deference to formal authority, a well-functioning bureaucracy, and effective enforcement all exist. Enforcement also depends on several practical considerations. The more resources regulators have, the greater competence of the regulators; the better the data, the simpler the process for detecting violations and imposing sanctions; the more voluntary the compliance, the better the police and court system—and the better the entire system of regulation will function to actually modify behavior. Simply put, enforcement of regulation is not the same in Brazil as it is in Singapore.

*Cultural differences***:** Large variations exist across nations in their general cultural attitudes toward the use of state coercion. Citizens of some societies are skeptical, or even anarchistic towards the government; other nations have a culture of trust in, or at least deference to, formal authority. Similarly, some governments are more willing than others to impose harsh penalties on violators, and violators know that if they are caught the sanctions will be severe. Overall, r*ules and sanctions that are seen as socially acceptable in a given culture are much easier to enforce*. High levels of voluntary compliance make the job of rule-enforcers that much easier, and the deterrent effect of enforcement that much more potent.

Category variation: Among the four categories of regulations, the purposes of each differ, the importance of each differ, the contributions to the final goals of a health system of each differ, and the enforcement difficulties of each differ. Establishing the basic conditions for market exchange such as specifying property rights, rules in buying and selling is easier to enforce since many victims of violators will complain and initiate remedial action. On the other hand, regulation has limited role and less effective in perfecting what the market can't do in *promoting equity.* Coerce private clinics and hospitals to give free services to the poor are unlikely to be complied. The state has to rely on another control knob, financing, to achieve its equity goals. Lastly, markets in

the health sector have serious failures. Consumers lack sufficient knowledge about the safety of many goods that can affect their health. Regulations to correct market failures in assuring the quality and safety of water, food, and pharmaceuticals have greater social acceptance and public support, and few alternatives are available to substitute state actions. Surely, preventing the public from the harms of carcinogenic chemicals or medical quacks is much more cost-effective in improving health. Regulation designed to correct market failures can have significant impacts on quality and efficiency of health services, cream-skimming the healthy people to insure and garner monopolistic profits for insurers and providers. The need to *correct market failures* depends on how much a nation relies on the market and for what. Also, the harms done by market failures are largely financial.

In sum, regulation, a control knob of the health system, can be an effective measure to affect the performance of a health system, particularly when it is combined appropriately with incentives, persuasion, and suitable organizational arrangement. Establishing regulation is relatively easy and inexpensive, but enforcement can be much more difficult and expensive. While regulation can significantly change behaviors in certain situations, but it can have insignificant impact in others. Health system reformers who want to make a real difference need to think carefully about the existing enforcement process, the resources available, and the incentives the regulation will generate, before going the regulatory route. All this adds to a simple point, good regulatory policy depends on good and wise judgment.

References

Abbott, Thomas A. III. (ed.) "Health Care Policy and Regulation." Boston/Dordrecht/London, Kluwer Academic Publishers 1995.

Ayres, I. And J. Braithwaite. Responsive Regulation, Transcending the Deregulation Debate, Oxford Socio-Legal Studies, Oxford University Press, New York, Oxford 1992, 39.

Bennett, S., Dakpallah, G., Garner, P., Gilson, L., Nittayaramphong, S., Zwi, A. (1994) Carrot and stick: state mechanisms to influence private provider behavior. *Health Policy and Planning*, **9**(1) 1-13

Bennett, S. and Ngalande-Banda, E., (1994) *Public and Private Roles in Health: A Review and Analysis of Experience in Sub-Saharan Africa.* World Health Organization SHS Paper Number 6. Geneva: WHO. **Contracting, and Regulation of Private Provider**

Bermann, George A. Managing regulatory rapprochement: institutional and procedural approaches, in: OECD, Regulatory Co-operation for an Interdependent World, Public Management Studies, Paris 1994, 73-91.

Bhat, R. (1996) Regulating the private health care sector: the case of the Indian Consumer Protection Act. *Health Policy and Planning*, **11**(3), 265-279

De Geynt, W. (1995) *Managing the Quality of Health Care in Developing Countries*. World Bank Technical Paper No 258. Washington, D.C.

Donaldbedian, A. (1980), *Explorations in Quality: Assessment and Monitoring, Volume I: The Definition of Quaity and Approaches to Its Assessment*. Ann Arbor, MI: Health Administration Press.

Fuller, Lon. The Morality of Law 28 (rev ed. 1969)—argued market economies require the establishment of property and contracts as institutions.

Heriot, Gail L. (1997) An Introduction: The Symposium on Law, Human Behavior and Evolution, 8 J. Contemp. Legal Issues I, 1-2

Herrin, A. (1997) Private Health Sector Performance and Regulation in the Philippines in William Newbradner (ed) *Private Health Sector Growth in Asia*, Chichester, England

John Wiley & Sons. Regulation of private insurance [103 private companies, assure performance of contract, or default. Solvency, risks taken, reserve, licnesing agents; Labor Code on provide health benefits to employees in company clinics, hazards in work place.

Hsiao, WC., Marketization *Health Economics.*

Hsiao, WC. Abnormal Economics

Jayasuriya, D.C. Regulation of Pharmaceuticals in Developing Countries, Legal Issues and Approaches, World Health Organization, Geneva 1985.

Kumaranayake, L. "Effective Regulation of Private Sector Health Service." Paper prepared for the World Bank Mediterranean Development Forum II, Morocco, September 1998. London, UK.

Laffont, J-J and J. Tirole. A Theory of Incentives in Procurement and Regulation. Cambridge (1993): MIT Press.

Leape, L., . To Err Is Human. Washington (2000), Academy of Sciences Press.

Nittayaramphong, S., Tangcharoensathien, V. (1994) Thailand: private health care out of control? Health Policy and Planning, (1), 31-40

North, Douglass C. (1990) Instituatioanl Chaange and Economic Performance pp. 69, 135-137; also Structure and Changes in Economic History (1981)—the need of legal protection of basic property and contractual rights for economic development.

Rule of law for human behavior and interaction. Oakeshott, Michael (1975) On Human Conduct.

Okma, K.G.H. Regulating the Private Insurance Market in The Netherlands. Rijswijk: Ministry of Health of the Netherlands 1995.

Roberts, Marc

Shkar, Judith, (1997) Political Theory and the Rule of Law in Hutchinson AC. And Monahan P. (eds) *The Rule of Law: Ideal Or Ideology 1.1*

Van Barneveld, E., van Vliet, R., van de Ven, W. (1996) Mandatory High-Risk Pooling: An Approach to Reducing Incentives for Cream Skimming. *Inquiry* **33**:133-143, Summer.

World Bank, (1997) World Development Report, 1997: The State in A Changing World.

Williamson, O. E. (1985) The Economics of Capitalism. Free Press

Chapter 11
Behavior

1. Introduction

Health system performance and health status are affected by individual behavior in multiple ways. For instance, sexual practices and needle sharing have major impacts on HIV/AIDS; whether people with tuberculosis take their medicines on the appropriate schedule affects the success of tuberculosis control programs; participation in vaccination campaigns influences the success of infant mortality reduction programs; physician prescribing habits for antibiotics influence the cost and the effectiveness of diarrheal disease control programs and the growth of microbial resistance; driving and seatbelt use affect the success of road safety campaigns. In short, changing individual behavior can have a major impact on both personal health status and health system performance.

As providers and patients, eaters and drinkers, drivers and lovers, individuals behave in various ways, in response to many forces. Some of these forces include the other control knobs: the financing and payment systems and the organization and regulatory structures. But behaviors are also grounded in culture and social structure, in habits, values, perceptions, beliefs, attitudes and ideas. This chapter describes the fifth control knob: methods for changing individual behavior through population-based interventions. We introduce the basic concepts of behavior change in relationship to health system performance and then explore the strengths and weaknesses of different behavior-change strategies for health improvement, with particular attention to health reform. We use the term *behavior* to indicate various approaches in the fields of information, education, and communication (IEC) and social marketing. Thus, the behavior control knob involves the design, implementation, and monitoring of programs intended to change individual behavior and influence the acceptability of social ideas, in order to improve health system performance.

Changing individual behavior often requires the coordinated application of the four other control knobs, in addition to messages aimed at individuals. For example, a behavior-change campaign could use monetary incentives (payment) to providers, or rules and sanctions (regulation). Similarly, the organization of the health care system can be altered to influence patient utilization of specific health facilities. In discussing the behavior control knob, we will refer to the four other control knobs where relevant. However, this chapter does not consider the use of behavior change strategies to influence the political acceptability of health policy proposals, since these issues are addressed in Chapter 5 on political analysis.

The behavior control knob draws heavily on the field of social marketing, and we use many of its concepts and strategies in this chapter. The field of social marketing began in 1969, with a seminal work by Philip Kotler and Sidney J. Levy (1969), who observed that marketing methods were being used throughout American society to sell everything from soap to colleges to politicians. They argued that all organizations confronting similar marketing challenges could benefit from marketing approaches,

whether the organization was selling a tangible commercial product or an intangible social idea. In this chapter, we have adapted the Kotler and Levy approach to social marketing for application to the issues of health reform.

Efforts to change individual behavior for public health goals occur within the context of pervasive private marketing, which seeks to shape behavior for commercial purposes. Private marketing invades all aspects of modern life, selling everything from cigarettes to coffee to computers to condoms to cough syrup. It explicitly emphasizes private business (and profits) more than public interests (and health), although some private marketing directed at meeting private health needs (such as condoms) can contribute to public health goals (such as control of sexually transmitted infections). Some private marketing has significant consequences for health status and health system performance, such as the marketing of food, pharmaceutical and tobacco products. Countries in transition from state-socialist to market-oriented systems have experienced a surge in commercial marketing, with dramatic changes in the patterns and amounts of private and public health expenditures. Sales of pharmaceutical products, for example, have exploded in the largely unregulated market of post socialist Vietnam (McManus, 1998).

Marketing by private medical practitioners and private hospitals also shapes public expectations about medical services and can lead to expanded purchases of high-technology medical equipment and services (such as CAT scanners) and expanded demands on tertiary hospitals, as has occurred in Korea (Yang, 1996). The behavior control knob provides a way for health policymakers to counterbalance the negative health consequences of private marketing and to redirect individual behaviors in ways that explicitly promote public health goals. The context of private marketing therefore provides motivations for using the behavior control knob, but it also creates political and practical challenges.

Efforts to improve health system performance by changing individual behavior can benefit from the lessons of commercial marketing. Behavior change efforts, like commercial marketing, must emphasize the central role of the customer and the core objective of meeting the customer's wants. Health reformers need to understand and work with the values, beliefs, and perceptions of the customer, which, after all, are the factors that shape health-related attitudes, knowledge, and behavior. Interventions that support or coincide with deeply held cultural values have a better chance of success, while interventions that challenge such values have a high risk of failure (as we noted similarly for regulation, in Chapter 9).

To be effective, health reformers need to combine different kinds of behavior change approaches. In general, reformers should not seek to "push" certain practices or ideas on people; instead, they should find ways to "pull" new, desired concepts to existing values held by the consumer.

The essence of this control knob is the voluntary decision by individuals to change their behavior – as opposed to regulation, which restricts behaviors through the

coercive power of the state. The degree of coercion in approaches to changing individual behavior thus varies from low-coercion approaches of simply providing information, to high-coercion efforts of restriction, indoctrination, and prohibition (Table 11.1).

Social marketing is more coercive than simply providing information, because it prepares, packages, and presents information with the intent of producing a specific change, but is less coercive than other approaches. The degree of coercion incorporated in behavior change interventions, however, can raise ethical concerns, which we discuss later in this chapter.

In designing behavior change strategies, health reformers confront three common marketing challenges (Kotler and Levy, 1969). First, they need a clear definition of the *product*, including both tangible products (including goods and services) and intangible products (such as ideas and beliefs). Second, they need to consider the product's relationship to specific *consumers* in the health system. Various kinds of consumers should be explored: the clients (or direct consumers of the product), the decision makers (or policymakers for the health system), key stakeholders (who have a specific interest in the product), and the general public (all those whose attitudes might affect the product's use in some way). Third, health reformers need to identify *tools* that can promote the product's acceptance by the consumers. In this chapter, we examine tools for changing health behaviors in relation to the *four Ps of marketing*: To develop "the right *product* backed by the right *promotion* and put in the right *place* at the right *price*" (Kotler and Zaltman, 1971).

The first applications of social marketing occurred in international development efforts in the 1960s. Family planning organizations in developing countries then began to employ social marketing in the late 1960s and early 1970s. At that time, contraceptive distribution in developing countries became a high priority for the US Agency for International Development and for private foundations, and they promoted the new methods of social marketing. The first national social marketing effort was the Nirodh Condom Project, which was initiated by the government of India, with assistance from the Ford Foundation (Altman and Piotrow, 1980). These projects commonly used commercial distribution channels, with government support and social marketing strategies to expand consumer demand. The projects also gave rise to many private consulting agencies that specialized in social marketing for family planning and health. In 1997, 60 large-scale contraceptive social marketing projects operated in 55 countries around the world. These projects sold a total of 937 million condoms and 54.5 million cycles of pills in 1997, and reported a 13% increase in number of couples served from the previous year – outcomes that have been attributed to the successful application of social marketing methods (DKT International, 1998).

Our primary concern in this chapter is how the behavior control knob can be used to improve health system performance and promote public health goals. To date, behavior change approaches have been directed mainly at particular health-related products: contraception, breastfeeding, tobacco, bed nets, immunization, and oral rehydration salts. These applications, however, have not been systematically connected to broader issues of

health system performance, and the behavior change approach has not been explicitly applied to implementing health sector reform in developing countries.[25] We argue in this chapter that many opportunities exist to influence individual behaviors within health reform plans, and that behavior change needs to be considered as a basic control knob for health reform, to improve overall performance and assist in achieving the core criteria.

2. Categories of Individual Behavior

Where can the behavior change control knob be applied in health reform efforts? Here it is important to consider four categories of individual behavior:

- treatment-seeking behaviors
- health professional behaviors
- patient compliance behaviors, and
- lifestyle behaviors.

Each category can be related to important objectives for health sector reform, including core criteria and intermediate criteria.

2.1. Treatment-seeking behaviors

Consumer decisions on when, where, and how to seek treatment represent an important area for improving health system performance. These treatment-seeking decisions include the kind of health professional (e.g., specialist versus generalist), the level of health facility (primary care versus teaching hospital), the timing and location of the treatment, and traditional versus modern medicine. These consumer behaviors often involve interaction with a health professional, including physicians, midwives, traditional healers, and private drug sellers. Treatment-seeking behaviors are important for acute health problems (e.g., trauma and fevers) as well as chronic health problems (e.g., diabetes and health disease).

One example of a behavior change campaign directed at treatment-seeking behavior would be an effort to encourage patients to visit the local primary care clinic before seeking care at the outpatient clinic of a university hospital, as a means to enhance the efficiency of the referral system (and to improve the intermediate criterion of cost-efficiency). Such a campaign might include the use of the slogan "Go First to Your Family Practitioner," along with efforts to improve the quality of services provided at primary care clinics (through the organization control knob). The effort could also use the payment control knob, by raising patient co-payments at the tertiary care facility and providing free treatment at the primary care facility. By using a gatekeeper mechanism (an example of the regulation control knob), the effort could also require patients to

[25] Health reform efforts have included the development of "communication plans," but these have been directed at the processes of *adopting* a reform proposal. We consider this use of marketing in chapter 5 on political processes.

receive approval from the primary care physician before being seen at the tertiary care facility.

2.2. Health professional behaviors

Provider decisions on the treatment for a patient or an illness represent a second important category of individual behavior that can be addressed through this control knob. Important provider decisions include the nature and frequency of the treatment, surgical versus medical versus pharmaceutical treatment, curative versus preventive approaches, the location of treatment (a public hospital versus the provider's private office), adherence to practice guidelines and treatment protocols, the referral of the patient to other providers, and in-patient versus out-patient versus home care. Questions about provider behaviors can include issues of professional ethics, and can be addressed by behavior change campaigns directed by professional societies and targeted at their members.

There is a huge literature on quality improvement for health services, frequently involving changes in provider behavior. One example in developing countries is the PROQUALI project, which seeks to improve the quality of reproductive health services in two states of Brazil. The project moves beyond conventional notions of clinical quality to a "new client-oriented model" of quality improvement. The PROQUALI approach involves organizational changes within the clinic, combined with external assessment for accreditation, followed by community recognition through an awards ceremony. The project also includes community mobilization, to enhance local demand for quality performance (PROQUALI, 2000).

Many behavior change strategies have been directed at the prescribing patterns of physicians and drug sellers. As pharmaceutical companies seek to promote the sales of their products in specific markets, drug prescribers are one of the main targets for commercial marketing in the health sector. Partly in response, the practice of "academic detailing" was developed, to change physician behavior in ways that would improve the quality and cost-effectiveness of prescribing (Soumerai and Avorn, 1990).

In developing countries, strategies to promote the rational use of drugs have included efforts to change the prescribing practices of not only physicians but also drug sellers, especially in environments where products are sold without a physician's prescription and often without the consultation of a trained pharmacist. For example, campaigns have targeted drug sellers to encourage the sale of oral rehydration salts and discourage the sale of antibiotics for cases of diarrhea (Ross-Degnan et al., 1996). Other campaigns have sought to discourage the use of injections, even though patients may want and expect this form of treatment as part of a medical visit.

These efforts illustrate a conflict that can arise between the core criteria of consumer satisfaction and health status. Injections may be the treatment that consumers prefer, but they may not be necessary or cost-effective, and can even cause health

problems through the reuse of needles that transmit infectious diseases such as hepatitis and HIV.

2.3. Patient compliance behaviors

The third category of individual behavior involves patient decisions on whether to follow the treatment instructions of health professionals. These behaviors include the use of pharmaceutical prescriptions, utilizing referrals to other health providers, and other kinds of treatment-following behaviors.

Many strategies aimed at changing patient compliance behaviors involve medications. Examples of behavior change strategies in this arena include efforts to encourage patients to take the full prescription of antibiotics (to decrease the development of antimicrobial resistance), and efforts to encourage patients with chronic diseases (such as schizophrenia, hypertension, and diabetes) to take their medications on a regular and timely basis. In some cases, social marketing strategies may be combined with direct regulation of individual behavior, to assure compliance, as occurs for tuberculosis, with directly observed treatment, short-course, known as DOTS (WHO, 1999). In directly observed therapy, health-care workers watch patients take their medications, to ensure that the right combination of drugs is taken for the right period of time.

Behavior change strategies have also been directed at persuading mothers to follow certain procedures in breastfeeding. For example, a social marketing program was introduced in traditional rural communities in The Gambia to combine traditional beliefs and modern knowledge in ways that would yield optimal breastfeeding practices, such as early initiation of breastfeeding, feeding of colostrum, and exclusive breastfeeding for six months (Semega-Janneh et al., 2001). The study recommended that attempts to change the breastfeeding habits of mothers need to take into account the attitudes and influence of husbands and elders, and especially the mothers of childbearing mothers.

2.4. Lifestyle behaviors

The fourth category of individual behaviors is consumer decisions on lifestyle habits that have major impacts on health status. Examples of these behaviors include individual decisions on exercise patterns, food consumption, smoking, sexual relations, and contraceptive usage. These behaviors do not necessarily involve interaction with a health professional; indeed, they are often subject to a complex set of influences, from modern commercial marketing to the force of traditional cultural expectations.

One example of a successful program to change lifestyle behaviors is the Harvard Alcohol Project's designated driver campaign (Winsten, 1994) as a means to change the social behavior of driving after drinking. The project persuaded Hollywood TV writers to incorporate references to *designated drivers* in top-rated television programs. These short messages within popular TV stories influenced people's perceptions of appropriate behaviors and provided social models of how to behave in specific situations. The project also found that friends and peers have critical influence on teenagers' behavior, so it

designed messages to connect to the existing values of teenagers, such as “Friends Don’t Let Friends Drive Drunk.”

The anti-smoking field has many examples of social marketing to change lifestyle behaviors, especially in the counter-advertising campaigns directed at the tobacco industry. Evaluations have shown several factors that raise the effectiveness of these campaigns. “Advertisements that directly attack the tobacco industry as the source of the tobacco problem; expose the way in which the industry manipulates, deceives, seduces, and addicts children and adolescents; and highlight the way the industry maintains adult smokers as lifelong drug addicts to make profits are effective in challenging the legitimacy and credibility of the industry” (Siegel, 1998, p. 130). Experiences from anti-smoking campaigns show how social marketing can be used in conjunction with regulation, payment, and financing to promote behavior change, and how evaluative research can help develop effective strategies that can be widely applied in different settings.

3. Basic Elements of Social Marketing

For health reformers, social marketing provides an effective conceptual framework for thinking about behavior change and developing practical strategies. Two other major approaches to behavior change are addressed by other control knobs: incentives (by the payment control knob), and restriction and prohibition (by the regulation control knob). Here we will use the social marketing framework to provide guidance to health reformers on concrete actions to change individual behavior as a means to improve health system performance. In this section, we present the basic elements of social marketing – organized around the four Ps of marketing: product, promotion, place, and price – as the critical dimensions that any individual behavior change plan must include. Health reformers need to consider these four aspects as part of their plans to change individual behavior. In exploring the strategies of social marketing, health reformers also need to consider interventions beyond media-based approaches, to include group interventions and other innovative approaches at changing individual behavior.

In addition, health reformers should design their efforts at behavior change based on a core assumption of social marketing: that changing a group’s basic values is extremely difficult; behavior-change strategies therefore should be built on existing values. This assumption serves as the foundation for commercial marketing, as well. Getting a population to adopt your product requires that you design the product to meet the values and needs of that population. This objective can be achieved by following the four Ps of marketing.

3.1. Product

The first step in social marketing is to define the product for a specific audience. This process is more complicated than it might seem, because it requires significant

analysis of different potential audiences. Three principles guide the definition and design of the product.

First is the principle of *consumer orientation*. Social marketing generates consumer satisfaction as a means of achieving social goals. To design an effective product, the marketer must know the consumer. Methods to know the consumer include quantitative methods (such as population-based surveys) as well as qualitative methods (such as focus groups). Information about the consumer is essential for product and message design, as well as for creating distribution and communication strategies. This kind of *formative* research helps define both the product and the audience for a social marketing effort.

This principle of consumer orientation provides the foundation for behavior change strategies. To be effective, efforts to change individual behavior should be based on a multidimensional understanding of the values and needs of recipients, and the product should be designed to meet (not change) those basic values and needs.

A second principle of product definition involves *audience segmentation*: The product should be designed for a specific population group. The broad potential audience for behavior change should be divided into separate target audiences, with an assessment of each segment's own interests, wants, and needs. Separate products should then be shaped for each market segment. For example, a Brazilian breastfeeding promotion program identified eight target audiences: 1) the doctor, 2) the health services, 3) the hospital, 4) the infant formula industry, 5) industry (in general), 6) the community, 7) government officials, and 8) the mother (Manoff, 1985, p. 48).

Audience segmentation is typically based on the psycho-social profile of a population (called "psychography" by marketers), identifying core values through an analysis of lifestyle, personality, communication habits, readiness for change, and perceived needs. These psycho-social characteristics can also be related to other factors that are used for audience segmentation, including "geography (region, county, census tract), demography (age, gender, family size, occupation, race, social class), and social structure (worksites, churches, voluntary agencies, families, legislative bodies)" (Lefebvre & Flora, 1988, p. 304). In the Brazilian breastfeeding program, each audience was defined, and a product was defined for each audience.

Qualitative and quantitative research of potential audiences can be used to create profiles of personality types for specific population segments. These profiles, such as the Traditionalist man in a low-income Filipino community, can then be used in designing social marketing messages for family planning (Piotrow et al., 1997, p. 42). Health reformers can use various techniques of consumer analysis to define the target audience, identify specific population segments, and describe the audience's core values. Market research and focus groups are important methods to understand how people think and feel about their health needs and health system, and to define relevant sub-groups in the population. Identifying the right audience segments is critical for defining products and

messages for each target audience, and for proposing specific methods to reach each group.

The third basic principle is to define the product as *meeting basic needs of the consumer* rather than simply providing a material object or a particular service. This principle also follows a commercial marketing tenet, a point made by Kotler and Levy (1969) in their classic article on social marketing. For a soap company, the product is not just soap but cleaning; for a cosmetics company, the product is not just makeup and lipstick, but beauty or hope; for a publishing company, the product is not just books, but information.

Similarly, to be effective, health reformers should view their product in functional terms as meeting basic needs and values. For a family planning organization, the product is not just condoms or contraceptives, but quality of life and reproductive control; for an AIDS prevention program, the product is not just safe sex practices, but freedom from disease and control over one's own destiny. Thus, the control knob for behavior change is concerned both with *tangible products* (such as condoms for contraception, or medications for chronic diseases) as well as *intangible products* (such as changing views about safe sex, or changing perceptions about quality of care). In short, health reformers need to learn how to define both kinds of products to appeal to the core values of a population segment in a specific society.

Table 11.2 provides some examples of how public health products can be connected to core values for particular segments of the population. For example, in considering smoking initiation by teenagers in developed countries, the core value of rebellion may be connected to decisions to start smoking (as adolescents rebel against their parents and their social norms). The challenge for social marketing is to connect the same core value of rebellion to a rejection of smoking (as an effort to rebel against the tobacco industry and its desire to hook teenagers on cigarettes). These core values need to be identified for population segments in specific countries, since values can vary across sub-populations and across cultures. A major challenge is to connect the desired changes to existing values in specific audiences.

3.2. Place

Once the products and audiences are defined, the social marketer next works on how to bring the product to the audience. Where can audiences be reached with the product? Making a decision on place involves the selection of appropriate *channels* through which the product can be delivered to the audience. As Lefebvre and Flora (1988, p. 305) noted,

> Any person, organization or institution having access to a definable population is a potential channel for health communication. Thus, schools, worksites, social organizations, churches, physicians' offices, and various nonprofit agencies can all be viewed as potential channels of communication. Identification of 'life path points' – such as laundromats,

grocers, restaurants, bus stops – can also uncover potential channels to reach certain audiences.

Health reformers need to think creatively about channels to reach the target audience, especially if the group is socially disadvantaged, as commonly occurs with public health campaigns for immunization, AIDS, or tuberculosis in developing countries.

Experience with social marketing shows that using multiple channels to reach your audience is most effective in changing individual behavior. Three broad categories of communication channels are commonly considered: *interpersonal* channels, including family, friends, and health care providers; *group* channels, including the mobilization community and civil society organizations; and *mass media* channels, including print, radio, and television (Piotrow et al., 1997, p. 73). Research has demonstrated that greater impact can be achieved by combining different kinds of channels, in order to repeat and reinforce key messages.

For example, a campaign in Kenya designed the slogan *Haki Yako* (It's Your Right) to emphasize the basic human right of women and men to control their own fertility, and then used multiple channels to disseminate that message: "radio spots and a radio serial drama to publicize the slogan; community visits by field workers to reach rural areas; posters and billboards to create a visual image of men and women talking together; and T-shirts to stimulate more interpersonal communication and to encourage advocacy by satisfied users" (Piotrow et al., 1997, pp. 73-74). A follow-up survey found that no single channel reached more than half the population, but the combined exposure from all five channels was 83 percent of a national sample survey. The goal for reformers is to reach target groups in different places through various channels, to raise the chances of changing knowledge, attitudes, and behavior.

The availability of diverse channels allows for imaginative approaches to connect the product and the audience. In using the behavior control knob, health reformers need to consider methods beyond conventional media approaches (such as radio), including advertising, personal contacts, social groups, and environmental changes (see Table 11.3). Specific audience segments have different characteristics and habits that affect where they can be reached. For example, public health promotion campaigns often use health centers to make available child survival products. But if there are problems in access and utilization of health centers – as commonly occurs in many developing countries – then using health centers for product distribution will not be very effective. Audience research for oral rehydration salts has identified other places as more effective for product promotion, such as private pharmacies (in Egypt and Indonesia), and small general stores (in rural areas of Honduras) – because the target population actually uses these facilities on a regular basis (Rasmuson et al., 1988, p. 11).

In some cases, special events can be organized to attract a particular population group, in order to deliver a message or a product. Social groups can also be used to promote and maintain behavior changes. For example, the Indonesian Family Planning Program (BKKBN) used village women's groups for promoting contraception and

organized village acceptor clubs to provide support through social and economic programs.

In the United States, the first statewide condom-distribution program used social marketing methods to decide on locations where condoms would be available (Cohen et al., 1999). The program selected health facilities (public health clinics, community mental health centers, and substance abuse treatment sites) to give authority and credibility to the role of condoms in disease prevention, through distribution by "a trusted health care provider."

In addition, the program provided free condoms in more than 1,000 private businesses in neighborhoods with high rates of sexually transmitted diseases and HIV, focusing on convenience stores, bars, nightclubs, liquor stores, beauty salons, barber shops, tattoo parlors, dry cleaners, and low-cost motels. The evaluation found an increase in self-reported condom use, without being associated with increases in the number of sex partners or decreases in private condom sales.

3.3. Price

Setting the price for behavior change interventions is a critical step that affects whether the product will be adopted and used. Deciding on the price involves managing both the monetary and non-monetary costs of product adoption, including social costs, time costs, geographic costs, and physical costs. These different costs can combine to create substantial barriers to the use of a product or service, or the adoption of a new idea or value. Applied research can be used to identify and measure these different costs for each product and audience and to find ways to overcome these costs to enhance the product's acceptance.

Price can also serve as a symbol of value to potential users. In some cases, a high price compared to similar products can attract purchasers who interpret price as reflective of high value. A low price can similarly lead potential users to avoid the product, out of concerns about low value. Setting the monetary price at zero, as a free product, can reduce financial barriers to access, but may result in unnecessary usage and waste. As noted in our discussion of the financing control knob (in chapter 7), economists are concerned that free products can encourage allocatively inefficient overuse. In other words, customers may use free health services even when the value to them is below the cost of production, indicating that more value could result if the resources were used to produce other goods or services in the economy.

A first step in setting the price for a social marketing effort is to examine the broader objectives of price policy. Nagle (1987) has identified five pricing objectives, all of which are relevant to different kinds of behavior change efforts.

3.3.1. Maximizing the number of product adopters

If the aim is to maximize the number of people to adopt a social product, then products and services can be offered at *low prices or for free*. Examples include free immunization for children, as is commonly done in immunization campaigns, and free prenatal care for pregnant women, as South Africa introduced in 1994 with the new government.

As noted above, however, free products or services can sometimes create the perception of low quality, which can discourage use, thereby reducing the level of product adoption. Alternatively, a low price can encourage overruse and waste, thereby creating inefficiency. Health reformers also need to think about the final price to consumers: subsidized price discounts may not reach consumers, since middlemen and retailers may sell the product at the normal price and pocket the discount as part of their profits (Manoff, 1985, p. 137).

3.3.2. Social equity

If a behavior change campaign seeks to achieve equitable distribution, then reformers can adopt a *graduated price structure* for the same product, with higher income groups paying more and low-income groups less. Examples of this pricing strategy include surgical procedures and the use of insurance premiums based on income level. Some community financing schemes also give an exemption in payment for the poorest members of the community. As mentioned in chapter 7, this kind of price discrimination is often used by medical practitioners who operate in a fee-for-service environment and charge high prices to rich patients and low prices to poor patients (thereby achieving increased access to poor patients and increased income to providers).

This pricing strategy can be used in behavior change efforts to reduce financial barriers to poor patients, while seeking to reduce subsidies to rich patients who can pay from their own personal resources or from insurance. Managing a graduated price structure, however, requires a method for means-testing (determining the economic status of users) to match the appropriate price with the customer, which can be a costly administrative procedure with high transaction costs. These problems can sometimes be addressed by innovative strategies for product targeting, such as offering low-priced products in particular geographic districts where rich customers are unlikely to go (as has been done for subsidized food products).

3.3.3.Cost recovery

If cost recovery from customers has a high priority, then a *fixed price* can be selected to defray a suitable part of the costs. An example of this pricing strategy is a flat fee for a provider visit at hospitals, regardless of the patient's income level, as a user fee to recover some of the operating costs. User fees (also discussed in the financing chapter) are considered a form of out-of-pocket payment, and represent a pricing strategy adopted in many health sector reforms, to allow public health facilities to recoup some of their costs directly from patients (in addition to government-provided resources).

The introduction of user fees, however, raises major questions about the equity consequences for poor patients, who can be discouraged from utilizing health services (reflecting the price elasticity of demand for different income groups). For example, the introduction of fees to treat bednets with insecticides (for malaria prevention) in The Gambia reduced utilization and increased mortality rates to levels before free treated nets were available (Armstrong Schellenberg et al., 2001). Social marketing efforts must be very careful not to set prices that can result in reduced access for poorer patients.

3.3.4. Demarketing

For certain types of products (such as tobacco and alcohol), social marketers deliberately set *high prices* in order to reduce undesirable or excessive demand. For example, this pricing strategy is used when a government imposes high taxes on alcohol and cigarettes, with the goal of reducing consumption, especially for price-sensitive groups such as adolescents. Cigarette companies, on the other hand, use their private marketing power to attract adolescents to their products, because this is the age when many smokers start a lifelong habit. High prices can also be used to reduce demand for medical services that are considered unnecessary or excessive, such as cosmetic surgery.

3.3.5. Profit maximization

While the pricing strategy of profit maximization is not commonly used for social products, one can imagine an organization that provides a social product, such as drug-abuse rehabilitation, and also seeks to maximize its profits (Kotler and Roberto, 1989, pp. 176-177). In that case, various price levels would be examined with a view to *how each price would affect demand and profits*. This pricing strategy is used most frequently by private companies. Multinational pharmaceutical companies, for example, set different prices for the same product in different countries, according to market conditions and income levels, in order to maximize overall profits.

Once the pricing objectives are clarified, social marketers need to decide on a price for the product. Three main methods are used for price-setting: an assessment of the costs involved, the prices for similar products from competitors, and the price-sensitivity of the target audience (Kotler and Roberto, 1989, p. 177). Each of these methods requires applied research, and the results must then be combined in ways that reflect the overall pricing objectives and goals of the behavior change endeavor. After the product's price has been determined, the health reformer can then turn to questions of promotion and communication.

3.4. Promotion

The promotion plan is designed according to decisions made about product, place, and price. The main objective of the promotion plan is to enhance the probability that consumers will accept the product, through messages that emphasize benefits and reduce costs. The promotion plan identifies specific activities and materials that will help achieve the campaign's overall goals.

The most important tasks in designing the promotion plan are to decide on the content of the message, how to present the message, and which channels of communication to use. To create an effective communication strategy that will reach the target audience and produce the desired change in individual behavior, health reformers need to bring together prior decisions about the product and the audience, the place to reach the audience, and the price of the product. Here we will discuss some of the critical elements of these three key decisions.

Social marketers typically develop multiple messages about a particular social product, in order to reach different audiences in the most effective ways. The content of the message is decided by combining information from four elements of the product (Kotler and Roberto, 1989, p. 225):

- physical/technical features (such as size, weight, shape, and other observable qualities)
- sensory features (such as qualities that people can feel, smell, see, hear, or taste)
- functional benefits (how the product will concretely help the target audience)
- emotional/psychological benefits (how the target audience perceives the product will help them)

Health reformers need to combine elements from these four areas to create the right message for a particular target audience. Experiences with anti-smoking campaigns show that the message needs to be empathetic and real, in order to persuade people to open up to the risks of behavior change. The message also needs to be believable, combining emotions and facts, and should illustrate the consequences of behavior change. Research on anti-smoking ads found that the most effective messages are ones that evoke strong negative emotions (fear and sadness). These ads focused on the consequences of smoking: "one [smoker] who had died, two who had lost their vocal cords to throat cancer, and one in which a baby who is exposed to secondhand smoke is coughing relentlessly" (Biener et al., 2000, p. 406).

The next task in promotion design is to decide how to present the message. Here the customer-orientation of marketing, which we discussed above under product design, is critical. Health reformers need to learn the basic rule of salesmanship: "Don't sell what you want, sell what they want" (Kotler and Roberto, 1989, p. 227). In designing the messages, marketers recommend a strategy of *emotion followed by logic*, or beginning with the product's benefits to the customer, and then moving to the product's features. Structuring the order of information in the message will depend on the specific target group and their responses to different kinds of information. Health reformers, therefore, need to know their consumers and how they react to information, images, and emotions.

The third step in the promotion plan is to decide on the channel of communication. Two main strategies are mass communications (to reach large target audiences) and personal communications (to reach individuals and small group audiences). Most social marketing programs use a combination of these two approaches, since they are mutually

reinforcing and more effective when used together (Kotler and Roberto, 1989, p. 190).[26] For mass communications, one of the frequently used channels is *advertising*, which gives the social marketer high control over content and timing of the messages. Budget constraints can reduce the effectiveness of this approach, although in some cases it may be possible to obtain free public service announcements.

In working with mass communications, it is important to develop good *media relations*, and by preparing materials and meeting with reporters, health reformers can frame the content of media reports. These efforts include preparing press releases (print, audio, and video), organizing press conferences, and conducting one-on-one briefings with key reporters. One way to obtain free media coverage is to organize a *social event* that can provide a 'hook' for reporters to write an article and thereby give public attention to a particular message. *Television and radio programs* can also be used to deliver health messages directed at individual behavior, such as the Designated Driver campaign in the United States, which was mentioned in many Hollywood television programs.

There are various methods of personal communication to reach individuals or small groups. *Social events* such as a sporting event can provide an occasion to reach a particular target audience with a product, such as safe sex practices at a soccer game. *Personal selling* by a peer or respected authority can be effective in reaching certain audiences. For example, it may be more effective to have a teenager speak with schoolmates on the issue of violence and about negotiation strategies than to have a schoolteacher do the same. *Entertainment* can be used to deliver health messages to particular groups, using both traditional and modern forms of entertainment, such as puppet shows that tour rural villages in India on the dangers of chewing tobacco. *Social groups* can also be used to provide a mechanism for both introducing and reinforcing new behaviors, and for assisting individuals with the transition to new habits. Examples of this approach include women's organizations that promote oral rehydration therapy or contraception, as well as Alcoholics Anonymous.

The final step in the promotion plan is to decide on actions that will *trigger* the adoption of the product. These trigger actions involve different kinds of incentives to persuade the customer to adopt the product (Kotler and Roberto, 1989, pp. 240-241), and can involve the payment as well as the organization control knobs. For example, incentives can be monetary (cash payment), social status (an award or seal of approval), or material (a radio or some other gift). One trigger technique is to provide a *free sample* of the product or the service, to induce the customer to try the product; another is to organize *contests and sweepstakes* for persons who agree to adopt the product, with the offer of prizes, including money, merchandise, or free services. A third trigger technique is to create *continuity programs*, in which adopters collect coupons each time they use a product or service, with a reward after a certain number of uses.

---

[26] A third approach of communication, known as selective communication, involves direct mail and telemarketing (Kotler and Roberto, pp. 212-220). This approach is probably not so relevant for health sector reform in most developing countries, due to factors of low literacy, sporadic mail service, and limited telephone availability.

One example of an entertainment-based promotion plan was the 1993 multimedia campaign to promote family planning in Mali (Kane et al., 1998). This project used traditional theatre and music forms, along with traditional proverbs in local languages and familiar settings, with broadcasts on radio and television. The objective was to change knowledge, attitudes, and practices related to contraception – to connect traditional social values with new contraceptive methods. The specific messages provided information about modern contraceptive methods, the need for male sexual responsibility, the health and economic advantages of family planning, the need for communication between spouses, and that Islam (the predominant religion of Mali) does not oppose family planning. A comparison of the situation before and after the intervention found increases in desired knowledge, attitudes, and self-reported practices along the lines sought by the social marketers. The authors concluded that the impact "may reasonably be considered to be both positive and significant" (p. 320).

4. Discussion

The main strength of the behavior control knob is its potential to address difficult public health problems based on individual behavior. These problems are common in health reform efforts and include the limited reach of public health services in developing countries for such services as immunization, prenatal care for pregnant women, and disease prevention for infectious as well as chronic diseases. Individual behavior patterns also shape the utilization of health services, including, for example, decisions to go directly to academic hospitals rather than use local health centers, decisions to depend on traditional healers or medicine sellers rather than seek care from licensed physicians, and decisions to avoid seeking out-patient care until a medical condition has become quite serious.

This control knob holds the promise of reducing or altering high-risk individual behaviors and of promoting protective behaviors, ranging from hand-washing to condom-using to helmet-wearing. This approach not only provides knowledge about health issues, but also connects that knowledge to core values in ways that encourage changes in individual behavior.

One example where behavior change approaches have been used to improve health system performance is the Delivery of Improved Services for Health (DISH) project in Uganda, supported by the U.S. Agency for International Development. The DISH project is focused on "improving quality, availability and utilization of reproductive, maternal and child health services, and enhancing public health attitudes, knowledge and practices in Uganda" (DISH, 2001) and uses many of the techniques discussed in this chapter to change behaviors that can contribute to improved health system performance, especially for the health of women and children. The project seeks to improve performance through a series of Best Practices, including the introduction of integrated reproductive health services, providing one-stop shopping in a client-oriented approach, and holding low-cost community events, which are recorded and broadcast as

radio shows, to create wider audiences for DISH messages. While the project does not include the entire health system in its objectives, DISH does provide multiple examples of how behavior change approaches can be successfully applied to achieve concrete improvements, especially in the core criteria of health status and customer satisfaction.

The examples of behavior change in this chapter illustrate that this control knob involves more than simply selling a product or redesigning a product for sale. The behavior control knob requires an iterative process of discovering the values of the target audience, creating goods and services consonant with those values, and evaluating the impact of implementing strategies designed to enhance the acceptability of the products. The example of motivating teenagers to practice being designated drivers shows how this approach can be effectively applied in practice.

Some public health policymakers may seek to use the behavior control knob to try to change the basic values of a target population, rather than focusing on trying to change behaviors. This contrasts with conventional commercial marketing, which typically seeks to develop and sell a product that will match existing social values. Private advertisers have learned how difficult it is to change people's basic values, leading the private sector to focus on identifying preferences and designing products to connect with existing values. When policymakers seek to change a community's basic values, they confront major challenges in achieving behavioral changes; indeed, trying to change basic values can undermine a campaign's ultimate effectiveness. It may be possible that a behavior change strategy can lead to broader social change, by changing the dominant patterns of behavior within a population (as has been argued for the birth control pill in American society). However, broad social changes often occur at times of multiple historical shifts, which make it difficult to draw clear causal linkages to specific individual behaviors.

Changing individual behaviors is critical to the successful implementation of health reform plans. The behavior control knob, therefore, needs to be used in conjunction with other health system changes in financing, payment, organization and regulation, in order to achieve the desired improvements in health system performance.

Behavior change, however, is not a magic wand. Social marketing methods present both problems and challenges, especially when compared to conventional commercial marketing. A list of some difficulties is presented in Table 11.4. Four points are worth discussing here.

First, public health professionals have a tendency to approach behavior change from an expert-driven or product-driven perspective ("We know what they need"), rather than seeking to understand and respond to consumer needs (Levebre & Flora, 1988, p. 302). This pattern reflects the bias of traditional medical training, without little emphasis on psychology and with a tendency toward arrogance. Furthermore, this expert-driven perspective conflicts with the basic consumer-oriented method of social marketing and creates a tendency not to conduct formative research for behavior change. It also tends to underestimate the importance of "pretesting … ideas, messages, and methods with representatives of the target group(s) *before* implementation" (Levebvre & Flora, 1988, p.

304). The requirement to *know the consumer* (with both quantitative and qualitative methods) cannot be overestimated in efforts to use behavior change in health reform.

A second point is that this control knob seeks to produce changes in individual behavior but does not seek to change the social structure or power dynamics of public health policy. The control knob focuses on changing individual behavior and personal responsibility rather than changing public policy. However, marketing methods are critical in public policy debates, in shaping public perceptions about the nature of social problems and appropriate policy solutions, and in influencing the acceptability of proposed policies; we deal with these issues in Chapter 5 on political strategies. Health reformers, therefore, should bear in mind the importance of conducting social marketing campaigns for changing individual behavior—in conjunction with applied political analysis and political mobilization for promoting social change.

A third point is that making behavior change work requires high-level commitment and substantial marketing expertise. The conditions when social marketing produces health behavior change have not been adequately specified (Walsh et al., 1993, pp. 115-116). In particular, we need additional studies on the kinds of behavior change approaches that work best for health reform and the conditions under which these approaches are most likely to succeed.

The final point involves the ethical aspects of using coercion to change individual behavior. How much coercion is considered justifiable depends in part on the nature of the problem and the values of the society. Coercion may be more acceptable for behaviors that involve harm to others (e.g., unprotected sexual behavior by a person who knows he is HIV-positive), that involve harm to self (for example, the decision by a motorcycle rider not to use a helmet), that raise cultural disapproval (for example, homosexuality or drinking in some societies), and that involve financial cost to society (for example, the health costs of illnesses due to tobacco smoking). Some societies are reluctant to use coercive methods for intimate behavior (such as condom use) compared to public behavior (such as seatbelt use).

The use of incentives to promote certain behavior changes (as discussed above for trigger activities in the promotion plan) raises a number of ethical questions about coercion. For example, what level of incentive in the promotion plan is considered coercive? Some people have argued that any use of incentives is coercive; others are concerned about the use of incentives for poor populations and for nonreversible procedures (such as sterilization). If the incentive works in persuading someone to adopt a behavior they would otherwise not do (such as a vasectomy, in order to receive a radio), then some would consider that offer coercive and ethically inappropriate. Health reformers need to decide how they will use incentives to promote behavior change, especially when directed at low-income groups.

The degree of coercion that is considered justifiable in changing individual behavior also depends on broader social ethics. The basic justification for coercive methods is objective utilitarian—that some individual behaviors are detrimental to the

maximization of health status for the population, and a higher level of coercion therefore is justified in seeking to change those behaviors. In short, the ends (better health) justify the means (coercion to change behavior). An objective utilitarian would favor the use of more coercive interventions, such as regulation and prohibition of specific behaviors, as well as quarantine (or incarceration) for recalcitrant individuals, as long as the approach would improve aggregate health.

On the other hand, a liberal would oppose any efforts to limit individual autonomy, even if respect for individuals resulted in unhealthy behaviors. But liberals would not object to the provision of information, as long as the autonomy of individual choice is respected. The problem is that the provision of information alone tends to be ineffective in changing ingrained personal behaviors important to public health (such as sexual practices, selection of health providers, water-use habits, and decisions about medication). In contrast to liberals, a communitarian would accept behavior change approaches to persuade providers and patients to comply with community values and would approve of using more coercive techniques for behaviors that violate cultural expectations.

These differences in the role of coercion, according to philosophical position, reflect one of the central ethical questions in public health: How much and what kinds of coercion are justified in seeking to change individual behavior for collective goals? This discussion of behavior change illustrates how health reformers need to understand the interactions of control knobs with each other – how behavior change requires simultaneous application of changes in financing, payment, regulation, and organization. It also shows the importance of ethical reflection and analysis for health reformers who use the behavior control knob.

Table 11.1
Approaches to Change Individual Behavior

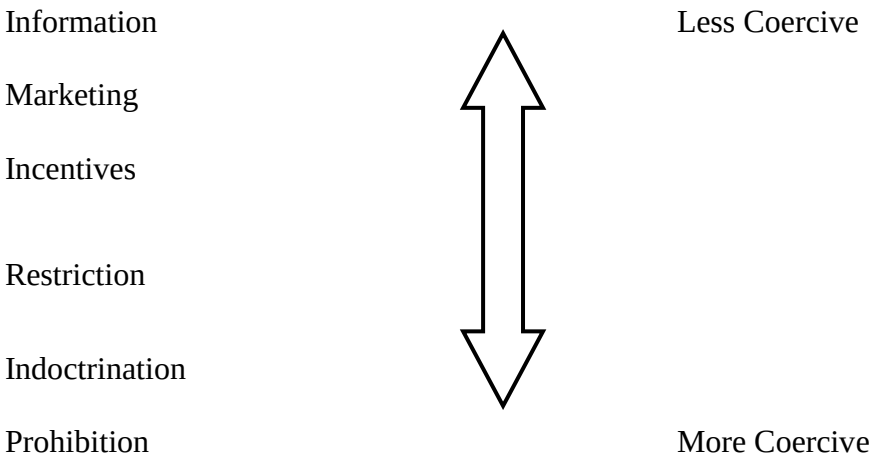


Table 11.2
Product, Benefit, and Core Values for the Desired Action Among a Target Audience

Desired Action	*Product/Benefits*	*Core Values*
Prevent smoking initiation	Freedom from nicotine addiction	Freedom
	Independence from tobacco industry manipulation	Independence Control
	Rebellion against an industry that is trying to trick you, seduce you, addict you, and kill you	Rebellion
Practice safe sex	Freedom from AIDS	Freedom
	Independence from the virus that is afflicting your friends and communities	Independence
	Control over your destiny	Control Rebellion
Exercise more often	Identity as a physically strong and attractive person in control of your appearance	Freedom Independence
	Rebellion against feelings of unattractiveness and lack of control over your appearance	Control Rebellion

Source: Siegel and Doner, 1998, p. 53.

Table 11.3
Channels for Reaching Target Audiences

- **Advertising**
- **Free media (eg, endorsements)**
- **Events**
- **Organized Groups**
- **Personal Contacts**
- **Entertainment**

Table 11.4

Problems and Challenges in Social Marketing

Market Analysis Problems

Social marketers

- Have limited primary data available about the consumers they are targetting.
- Have difficulty obtaining valid, reliable measures of salient variables.
- Have difficulty sorting out the relative influence of identified determinants of consumer behavior.
- Have difficulty getting consumer research studies funded, approved, and completed in a timely fashion.

Market Segmentation Problems

Social marketers

- Face pressure against audience segmentation, especially if segmentation leads to ignoring certain segments.
- Frequently do not have accurate behavioral data to use in identifying segments.
- Often target consumers who are negatively predisposed to the proposed changes.

Product Strategy Problems

Social marketers

- Tend to have limited flexibility in shaping their products or offerings.
- Have difficulty in formulating product concepts.
- Have difficulty selecting and implementing long-term positioning strategies.

Pricing Strategy Problems

Social marketers

- Find that the development of a pricing strategy primarily involves trying to reduce the monetary, psychic, energy, and time costs incurred by consumers when engaging in a desired social behavior.
- Have difficulty measuring their prices.
- Tend to have limited control over consumer costs.

Channel Strategy Problems

- Social marketers have difficulty utilizing and controlling desired intermediaries.

Communications Strategy Problems

Social marketers

- Usually find paid advertising impossible to use.
- Often face pressure not to use certain types of appeals in their messages.
- Usually must communicate relatively large amounts of information in their messages.
- Have difficulty conducting meaningful pretests of messages.

Organizational Design and Planning Problems

Social marketers

- Often work in organizations where marketing activities are poorly understood, weakly appreciated, and inappropriately located.
- Often function in organizations where plans (if any are developed) are treated as archival rather than action documents.
- Often function in organizations that suffer from institutional amnesia.

Evaluation Problems

Social marketers

- Frequently face difficulties trying to define effectiveness measures.
- Often find it difficult to estimate the contribution their marketing program has made toward the achievement of certain objectives.

Source: Bloom and Novelli, 1981.

References

Altman, Diana L., and Phyllis T. Piotrow, "Social Marketing: Does it Work?" *Population Reports* J-21, January 1980.

Armstrong Schellenberg, Joanna R.M., Salim Abdulla, Rose Nathan, et al., "Effect of Large-Scale Social Marketing of Insecticide-Treated Nets on Child Survival in Rural Tanzania," *The Lancet* 347:1241-1247, 2001.

Biener, Lois, Garth McCallum-Keeler, and Amy L. Nyman, "Adults' Response to Massachusetts Anti-Tobacco Television Advertisements: Impact of Viewer and Advertisement Characteristics," *Tobacco Control* 9:401-407, 2000.

Bloom, Paul N., and William D. Novelli, "Problems and Challenges in Social Marketing," *Journal of Marketing* 45:79-88, 1981.

Cohen, Deborah A., Thomas A. Farley, Jean Roger Bedimo-Etame, Richard Scribner, William Ward, Carl Kendall, and Janet Rice, "Implementation of Condom Social Marketing in Louisiana, 1993 to 1996," *American Journal of Public Health* 89:204-208, 1999.

DKT International, *1997 Contraceptive Social Marketing Statistics*, Washington, DC: DKT International, 1998.

Horsburgh, C. Robert Jr., "The Global Problem of Multidrug-Resistant Tuberculosis: The Genie is out of the Bottle," *JAMA* 283:2575-2576, 2000.

Kane, Thomas T., Mohamadou Gueye, Ilene Speizer, Sara Pacque-Margolis, and Danielle Baron, "The Impact of a Family Planning Multimedia Campaign in Bamako, Mali," *Studies in Family Planning* 29(3):309-323, 1998.

Kotler, Philip, and Sidney J. Levy, "Broadening the Concept of Marketing," *Journal of Marketing* 33:10-15, 1969.

Kotler, Philip, and Gerald Zaltman, "Social Marketing: An Approach to Planned Social Change," *Journal of Marketing* 35:3-12, 1971.

Lefebvre, R. Craig, and June A. Flora, "Social Marketing and Public Health Intervention," *Health Education Quarterly* 15(3):299-315, 1988.

Manoff, Richard K. *Social Marketing: New Imperative for Public Health*. New York: Praeger, 1985.

McManus, Joanne. *The Future of the Pharmaceutical Industry in Asia*. New York: Economist Intelligence Unit, 1998.

Nagle, Thomas T., *The Strategy and Tactics of Pricing*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

Piotrow, Phyllis Tilson, D. Lawrence Kincaid, Jose G. Rimon II, and Ward Rinehart, *Health Communication: Lessons from Family Planning and Reproductive Health*. Westport, CT: Praeger Publishers, 1997.

"PROQUALI Improves Health Services in Brazil," *Communication Impact!* No. 10, August 2000.

Rasmuson, Mark R., Renata E. Seidel, William A. Smith, Elizabeth Mills Booth, *Communication for Child Survival*. Washington, D.C.: Healthcom, June 1988.

Ross-Degnan, D., S.B. Soumerai, P.K. Goel, et al., "The Impact of Face-to-Face Educational Outreach on Diarrhoea Treatment in Pharmacies," *Health Policy and Planning* 11:308-318, 1996.

Semega-Janneh, Isatou Jallow, Erik Bohler, Halvor Holm, et al., "Promoting Breastfeeding in Rural Gambia: Combining Traditional and Modern Knowledge," *Health Policy and Planning* 16:199-205, 2001.

Siegel, Michael, "Mass Media Antismoking Campaigns: A Powerful Tool for Health Promotion," *Annals of Internal Medicine* 129: 128-132, 1998.

Siegel, Michael, and Lynne Doner, *Marketing Public Health: Strategies to Promote Social Change*, Gaithersburg, MD, Aspen Publishers, 1998.

Soumerai S.B., and J. Avorn, "Principles of Educational Outreach ('Academic Detailing') to Improve Clinical Decision Making, *JAMA* 263:549-556, 1990.

Wallack, Lawrence, and Lori Dorfman, "Media Advocacy: A Strategy for Advancing Policy and Promoting Health," *Health Education Quarterly* 23(3):293-317, 1996.

Walsh, Diana Chapman, Rima E. Rudd, Barbara A. Moeykens, and Thomas W. Moloney, "Social Marketing for Public Health," *Health Affairs* (summer):104-119, 1993.

Winsten, J.A., "Promoting Designated Drivers: The Harvard Alcohol Project," *American Journal of Preventive Medicine* 10(3 Suppl):11-14, 1994.

World Health Organization, *What is DOTS? A Guide to Understanding the WHO-Recommended TB Control Strategy Known as DOTS*. Geneva: WHO, 1999.

Yang, Bong-min, "The Role of Health Insurance in the Growth of the Private Health Sector in Korea," *International Journal of Health Planning and Management* 11:231-252, 1996.

Chapter 12
Conclusions

We have tried, in the preceding chapters, to provide readers with a set of tools and concepts that will help them conduct health sector reform in their own countries more effectively. We have avoided presenting a set of "cook book" recommendations. Instead we have tried to provide more general guidance on the art and science of cooking – so readers can develop and use their own recipes, ones that suit their own local conditions and food preferences. We have tried throughout to be transparent about our own views – philosophical and scientific. This is so readers can calibrate our presentation and add a "grain of salt" where they find it necessary.

We have also tried to be honest about the difficulties of the health sector reform - process. Different interest groups, political parties and bureaucratic agencies will have different priorities and favor different policies. Entrenched forces will resist change. The complexities of the system mean that almost any policies can and will produce unintended consequences. And events in the wider world – ranging from economic instability and political turmoil to social and cultural developments and climatic cycles – will all influence what seems either possible or necessary.

Yet we also believe that it is possible to do things differently – and better – than they are normally done. Honest and self-critical analysis can help would-be reformers "get health sector reform right." The concepts we have offered: the policy cycle, the ethical framework, the performance criteria, political mapping, diagnostic trees, and the control knobs are all designed to be used in that way. They are only a beginning, and must be applied in a context-sensitive way. But they are a beginning: a framework for organizing work, doing analysis and generating and evaluating options. Obviously they are not self-implementing – but we hope and believe they can be helpful.

The analysis we have presented, explicitly and implicitly depends on series of critical propositions about the structure and functioning of health care systems, about how health sector reform processes typically work themselves out, and about how reformers ought to view their own responsibilities. We want to first present these propositions and then discuss their implications for the reform process.

- *Health systems are complex socio-economic entities, but their structure and functioning can be influenced by rational action.* These systems in any one country has evolved over time as the result of an evolving mix of explicit public policy, multiple private actions, history and culture. But continuing research in recent years has begun to clarify how that structure can be changed and what difference various changes are likely to make. We have summarized the impact of these variables under the headings of our five "control knobs," and this emerging body of evidence constitutes the real scientific basis for health sector reform.

- *Various features of the health system have evolved to suit different, often conflicting objectives.* The complex and historical embeddedness of health sector arrangements often produces inconsistent policies and self-contradictory structures. Like the varying rock layers underlying a geological formation, different eras of political action have often been the product of varying configurations of political forces and moved the system in different directions. Different interest groups may have captured various institutions and parts of the system and bent them to their respective selfish purposes. It should not be surprising, then, that the system does not aggressively and effectively maximize some simple, specific objective function.

- *Reformers can only conduct the process of reform effectively if they can specify what goals and objectives they are trying to reach*. Since the goals of the system are both numerous and subject to disagreement, reformers cannot see themselves as mere technocrats. If they don't recognize the critical role of values, they will not be able to solve even the simplest technical problem, like setting priorities. There is no way to do that without, for example, deciding on the relative importance of saving citizens at different ages. Reform is a value-relative process, through and through.

- *Changing the healthcare system will inevitably be a political process*. Rational analysis might tell reformers how best to achieve their goals, in a technical sense. But that is not enough to produce change in a system whose roots and structure is so deeply political. Whether the government is democratic, monarchical or run by a single party – groups must be mobilized through the policy cycle if that cycle is to proceed. And that process of mobilization is at its heart political.

- *Designing an effective reform package will require both a deep understanding of the system and a willingness to act on that understanding.* We have stressed repeatedly that the health system can be understood. But achieving that understanding is not always easy. And acting on such insight may be even more difficult. Those adversely affected will resist change – and they often have substantial resources with which to do so. The diagnostic process we have described is thus critical to successful reform initiatives.

We do not, however, want to leave the reader with the impression "It all depends." For that, in fact, doesn't say much about what depends on what. Thus we do want to offer summary pieces of advice out of our control knob discussion.

- *No financing system can produce more resources than a nation's economic system generates.* Nations both tend to, and should, adjust their financing arrangements to their level of economic, social and political development. But they have to accept that the resources they have to spend on health care are limited by that level of development.

- *Because resources are always limited, nations need to adjust payment, organization and regulation to increase technical and allocative efficiency.* The incentives on organizations, generated by payment and regulation, and the capacity of organizations to respond – as a result of their micro and macro structure – need to be adjusted

consistently. This is the only way to maximize clinical and service quality and to assure the right mix of outputs. If a society is to maximize both health status and satisfaction it has to be willing to employ settings on these control knobs designed to achieve these results. This means increasing incentives on and the accountability of organizations, and increasing their capacity to respond to these incentives.

- *Failure to engage in reforms designed to improve outcomes – however understandable politically – will only perpetuate poor performance*. Many in the health care system do *not* in fact want to run the system to improve its performance. Instead they want to defend their incomes, or maximize political benefits, or provide gains to their local constituents. While such motives are understandable, they help insure the perpetuation of a system that under-performs relative to performance criteria.

- *Think about both the "demand and "supply" side consistently*. There are limits to how much any state can coerce citizens to use health services they find unappealing or unattractive. The supply of services – if it is to be utilized – has to be organized and priced in a way that encourages utilization. Complementary efforts to influence individual behavior likewise have to be focussed at responding to and fulfilling citizens underlying needs – rather than trying to reshape them to fit some bureaucratic or technocratic ideal.

Finally we would like to offer a summary set of injunctions for reformers to bear in mind. In keeping with our whole purpose – our conclusions are stated in the form of *advice* to those who would engage in health sector reform.

1. Clarify your goals

Health sector reform is not about solving a single, well defined, set of problems. Instead it is a messy, contentious process – often initiated in response to some external shock – about which there will be much disagreement. Therefore if reformers are going to improve their national system's performance in some way or other, they are well advised to begin by clarifying their goals.

We have stressed throughout that the health system has to be seen as a means to an end – and have offered the intermediate and final performance criteria (Chapter 4) as a way of describing those ends. And we have stressed that different ethical theories (Chapter 3) – as well as differences in politics and interests – will lead to different goals. Maintaining services to the poor is much more important to egalitarian liberal societies like Sweden or Denmark than it is to the far more libertarian U.S. By the same token Singapore is much more willing to coerce its citizens – based on expert driven analysis of what will improve their health – than would be acceptable in America.

Clarifying goals is especially important because goals can conflict and different goals can imply different policies. Concentrating high-tech services in a national facility (like the National Heart Institute in Kuala Lumpur) can take advantage of economies of

scale – but it will also reduce competition. Charging user fees for hospital use in Uganda will improve efficiency by discouraging low valued usage – but also tend to exclude the poor, leading to equity issues. Trying to lower IV drug use via aggressive policing might or might not be effective in changing behavior in ways that diminish the spread of HIV – but it is also likely to lead to a decline in personal freedom.

Since the cycle of health sector reform is often initiated in response to an external shock, reformers have to be aware that certain definitions of "the problem" are likely to be current in the media and in political circles. These will often be promoted by specific economic or bureaucratic interests. The doctors will agitate for higher incomes, the social insurance fund will ask for increased subsidies, the Ministry of Finance will point to existing waste and inefficiency.

Rather than take any of these definitions as definitive, however, reformers need to take responsibility for the priorities they set and the goals they identify. This means they need to openly articulate, and advocate for, their perspective in the messy, hurly-burly world of political debate. We have argued that they can best do this – from the perspective of clarity and coherence – by relating these performance goals to their own underlying ethical position.

Choosing priorities, we have argued, is best done strategically. Problems that emerge due to ethical concerns should be subject to a *strategic analysis* focused on political feasibility and the likely availability of policies that will have a positive impact. And in doing this, various kinds of benchmarking can help set your aspiration levels.

2. Do an Honest Diagnosis

Once priorities have been identified and performance problems described, reformers next have to do an honest diagnosis. We strongly urge the drawing of multiple "diagnostic trees" to reveal patterns of causality and points of critical importance to the system. This is the practical method for giving life to the advice to "Ask 'why' five times."

We have stressed the need for honest diagnosis because official reports and discussions often find it difficult to confront certain "taboo" features of the system. Widespread unofficial payments to doctors, endemic hiring of the politically well connected, physicians in rural areas who don't even take up their posts, massive theft of hospital supplies, performance reports full of fictitious numbers, poor health status among marginalized groups; these are just some of the kinds of issues and patterns that diagnostic journeys need to honestly address. In this connection, international experts and/or those at home with a variety of positions and perspectives can help challenge existing assumptions and reveal underlying premises.

We also recognize however that too much time and resources spent on analysis can delay action and paralyze reform efforts. To accelerate the process, try to use existing studies

and existing data and not feel a need to do everything again. They also need to know when to stop. In the real world data is always imperfect and effective reformers need to master the art of balancing better results against added costs. Given the cyclical nature of health sector reform, the window of opportunity for mobilizing action will not remain opened indefinitely.

A key aspect of our diagnostic advice is to *not* begin with one's favorite solution – or even with one's favorite target for abuse. Instead we urge an "evidence based" approach – relying on data and science as opposed to slogans or pre-conceptions.

3. Develop a Plan That Can Reasonably Be Expected to Work in Your National Context

We have stressed repeatedly that solutions to national problems must fit the local context (see, for example, the diagram summarizing health care financing option in Chapter 7). Moreover this is a matter of cultural, political and institutional "fit" as well as correlation to the level of economic development. If bureaucrats cannot be trusted not to steal the money in the social insurance fund (as they did in Kazakhstan), or if firms will evade payroll taxes by shifting to non-cash compensation to employees (as they did in Hungary) then such possibilities need to be considered when programs are designed.

Although we have organized our discussion of interventions in terms of distinct "control knobs," we have stressed repeatedly that effective policy may require coordinated action across these conceptual boundaries. Changing payment systems to hospitals, to create incentives for efficiency, will do little good if hospital administrators, at the micro-organizational level, are not also provided with the skills, incentives and authority to respond. A social marketing campaign to encourage use of local primary health care centers is less likely to be successful if, at the same time, we raise patient co-payments for services at those centers.

We have also stressed that even within a control knob, consistent policy may be required across a number of interventions. Few nations rely on only one financing mechanism for example. And if we want to use payment schemes to create incentives to avoid inefficient care, we may need to line up both how doctors are paid and how insurance is paid for.

In designing policy, we have repeatedly urged realism about what is possible. For example people in the middle of the income distribution will pay typical health care costs directly or indirectly. A good financing system therefore needs to focus on providing risk protection, (especially for the vulnerable) rather than on the unattainable goal of avoiding the overall cost burden. Similarly if local political leaders have a long record of using any resources and authority decentralized to them for patronage purposes, we cannot assume that health sector decentralization will improve customer service.

The art of effective reform requires a balance between being overly ambitious and being overly pessimistic. Both kinds of mistakes are possible. Trying to design a "perfect" plan – one that anticipates all contingencies and problems – is likely to take too long and not succeed anyway because of system complexity. Instead it may be better to get going and amend the details incrementally – as Singapore has done with its medical savings account system. At the same time, seeing only the obstacles and difficulties can lead reformers to do too little. Pushing the envelope may be a better strategy – again with the chance to amend and adjust in later policy cycles.

We have argued again and again for the value of international learning. And the control knob chapters in Part II are full of examples – both positive and negative. But in using such experience we caution again – be evidence based not ideology based. Consider your own national context.

Finally reformers need to recognize that the process by which they develop their reforms will have much to do with their acceptability. Participation of key actors and interest groups can produce valuable input, and help involve and co-opt potential opponents. On the other hand, avoid any commitment to unanimity. It is not possible to both do genuine reform and please everyone. Negotiate? Yes. Turn the "asylum" over to the "inmates"? No.

Reformers do have to consider how the inmates (consulted or not) are going to react if and when their policies are implemented. Especially when it comes to financing and regulation, the coercive power of the state is being deployed. And those being coerced are likely to resist. Designing policy that is seen as legitimate, using data that is inexpensive and reliable, lowering the costs of collection and enforcement, organizing the relevant agencies so they have the right skills and incentives – all of these are key matters to consider to insure the implementability of whatever policies are being introduced.

4. Embrace Politics

Reform, we have argued since Chapter 1, is not just a technical but also a political matter. Politics pervades all stages of the reform cycle. This means reformers need to embrace, not shun politics. They need to do the kind of explicit political mapping and strategy development discussed in Chapter 5. Exactly because there are differences in ethics and values – not to mention in interests and in beliefs about how the world works – reformers cannot expect that everyone will agree. And political processes of one sort or another will inevitably be deployed to resolve these disagreements.

The fact that the health reform cycle is a cycle also means the politics of reform will continue to evolve. Reform will strengthen some interest groups and weaken others. New external shocks will occur. Just as reformers need to be prepared to re-evaluate their substantive analyses – they need to re-do their political analyses also.

We have stressed repeatedly that successful reform is not just a matter of "the merits." Instead it is also a matter of symbolism and language, of political skill and personal commitment. While we have urged realism we also are aware that reformers can all too easily conclude that change is impossible – or at least unlikely. The forces defending the status quo are often so impressive – organized, activated and committed. But reform does happen. Columbia does have a new insurance system. Korea is changing the integrated role of doctors and pharmacists. Hungary is paying hospitals differently. These successes have been achieved by reformers who are willing to use political processes shrewdly to advance their case.

5. Focus on Implementation

Successful implementation often requires leadership and attention to detail. If local authorities are to get more responsibility, they may need intensive training first (as they did in Kerela). If fiscal decentralization to improve inter-regional equity (as it did in Chile) as opposed to undermine that goal (as it did in Bosnia) how inter-regional re-allocation systems need to be carefully defined.

Successful implementation requires performance measurement. But here again realism is critical. Some districts in India report 120% immunization coverage (defined as doses delivered divided by the eligible population). But a closer look revealed that 40% of the children were getting, on average, 3 doses each! Lean too hard on a measurement system and you risk dishonesty and/or unintended distortions in behavior. In the U.S., clinics that do I.V.F. report their success rates in a standardized way that is reliable—but few will do "difficult" cases to avoid pulling down their reported outcomes. Again we come back to the need to focus on performance, on outcomes not inputs, on results not on efforts made.

We have noted repeatedly that institutions resist change out of both interest and anxiety. Soshka's research in Hungary calls attention to the fact that managers can lower some of that resistance by lowering uncertainty. Staffs need to be helped through the reform process by leaders who are both credible and committed, and who understand the dynamics of organizational change.

Implementation won't just happen. It needs to be planned for. Critical tasks need to be identified and likely obstacles anticipated—which of course they cannot be fully. Managers need to adapt as experience accumulates. Many reformers, especially at senior (Ministerial) levels turn over so quickly that they are not in place long enough to really see to implementation. Such transience is hardly likely to help produces measured improvements in health sector performance.

6. Learn from Your Mistakes

Even successful reform will all-too-often only lead to new problems. Reformers therefore need to be prepared to learn from their mistakes. This requires us to build evaluation systems *into* the very process of policy design. They need to create incentives for providers and insurers to report data accurately. They need to consider experimental design and collect baseline information before beginning. They need to expect that some of the details (and maybe even some of the major features) of the system will be wrongly specified the first time.

This is why the policy cycle is cycle. This is why reform is not a one-time or one-off activity. The large-scale forces we reviewed in Chapter 1 will continue to operate. Technology will continue to advance. Expectations will increase. Costs will go up. Economic turmoil will occur. Political coalitions will be unmade and remade. Our hope is that reformers will commit to the process for the long run. That they will think critically and politically, that they will be self-reflective ethically and creative institutionally, that they will focus on performance and on policy details. All this is quite demanding, we know. But it is what is required to do a better job of "getting health reform right."

wb231172
M:\hadia\Harvard Book\Master.doc
October 24, 2001 1:35 PM