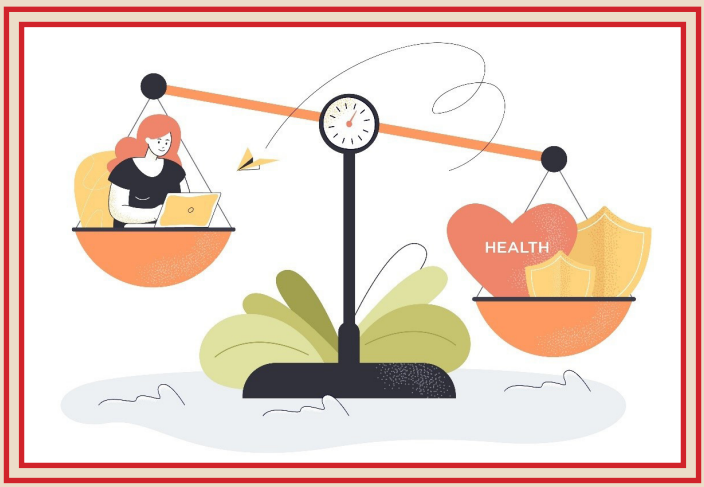
BASICS OF ECONOMICS IN HEALTHCARE

LECTURE NOTES



MINISTRY OF HEALTH OF UKRAINE
NATIONAL UNIVERSITY OF PHARMACY

A. S. Nemchenko, N. V. Demchenko

BASICS OF ECONOMICS IN HEALTHCARE. LECTURE NOTES

STUDY GUIDE

Independent electronic educational publication

Kharkiv
National University of Pharmacy
2024

UDC 339.9
N 39

Reviewers:

D. V. Karamyshev, Doctor of Science in Public Administration, Professor of the Department of Disaster Medicine and Military Medicine, Poltava State Medical University;

L. O. Gala, Doctor of Pharmacy, Professor of the Department of Organization and Economics of Pharmacy, Bogomolets National Medical University.

Recommended
by the Academic Council of the National University of Pharmacy
(protocol No. 11 from 11.12.2023)

Nemchenko A. S., Demchenko N. V.

N 39 Basics of Economics in Healthcare. Lecture notes : study guide. – Independent electronic edition. – Kharkiv : National University of Pharmacy, 2024. – 154 p.

ISBN 978–966–615–613–9 (PDF)

The lecture notes contain the main topics of the educational component "Basics of Economics in Healthcare" explaining the basic economic categories, laws, and regularities, principles of organization and construction of economic relations, covering issues of pharmaceutical and medical services, economic assessment in healthcare, etc. The lecture notes have a convenient and logical structure and illustrations.

For classroom and extracurricular training of English-speaking students majoring in specialty 226 "Pharmacy, Industrial Pharmacy" at the first (bachelor's) level of higher education.

UDC 339.9

ISBN 978–966–615–613–9 (PDF)

INTRODUCTION

Economics in Healthcare is a fascinating subject. This course is designed to familiarize students studying pharmacy and economics with the basic economic principles and their application to the healthcare sector. Hence, this course material should be regarded as an introduction to health economics rather than to economics itself. The overall purpose of the course is to introduce the basic concepts of economics and their application to the health sector and rather than to fully present everything that is important about the subject of health economics. Thus, the need for additional reference books may be of paramount importance. The concepts and analyses presented in this course material will help to serve as working material so that students and other professionals can understand and apply the basic ideas of economics to the health sector. The topics covered include special features of the healthcare market, the four basic issues, the main characteristics of the healthcare service and its relationship to the economic development, economic models and analysis, the health insurance market, the issue of equity as it relates to health and healthcare, the methodology of health economic assessment and the economics of healthcare and sustainable development.

The aims of this course are to provide in-depth understanding of health economics in terms of:

– values and goals of health economics;
– the main tools for economic analysis of the health sector;
– healthcare as one of the social sectors of economic importance;
– the specific nature of the healthcare service;

– the importance of economics to resource allocation, planning and management of the health sector;

– the impact of the economic development on the health services;

– economics of healthcare financing.

To achieve the aims of this course, there are general and clearly defined objectives which the course is set to achieve for each unit. The unit objectives are indicated at the beginning of the unit; you should read them before starting work on the unit. You may want to refer to them while studying the unit to assess your progress. You should always look at the unit objectives after completing the unit. This is done in order to help students complete the tasks provided for in this course.

UNIT 1

THE ECONOMIC BASIS OF HEALTHCARE, ITS TASKS, AND FEATURES IN A MARKET ECONOMY

Content

Introduction

Economists in all sectors are concerned about the allocation of resources between competing needs. It is assumed that the needs are infinite – there is no end to consumer aspirations. Resources such as labor, raw material, production equipment and land are always finite. Thus, scarcity of resources becomes the fundamental problem to which economists address themselves. In the health-care sector, such scarcity can be recognized in the multitude of issues that concern everyone who works there or uses its services. Why has the volume of resources absorbed by the sector increased so fast over the last four decades worldwide? Why does it seem that no matter how many nurses and doctors are employed, new technologies adopted, and new drug therapies introduced, even the rich countries of the world are not able to provide the highest quality of care for all citizens? Why do economists work in the healthcare sector? The healthcare sector is not the first place people

associate with economists. In principle, economists are concerned with better choices and in making the best use of existing resources and growth in the availability of resources. As economists started to work on problems in the healthcare sector, the new discipline of health economics emerged. Many of the concerns in health economics are also those of other health scientists – how can we improve survival, quality of life, and fairness in access to services? However, economics brings a different framework that offers important and useful insights for analyzing such questions. Therefore, understanding the modern economy requires an appreciation of the special economics of healthcare.

This module starts by examining various views on the definition of health economics within the various schools of thought in economics. It also examines various economic problems in the healthcare sector. It looks at the special features of the healthcare sector and the four basic questions on the allocation and distribution of resources in the healthcare sector. This module further examines the measurement of the health status and possible future challenges of the healthcare sector. The proper scope of government intervention in the healthcare system is a topic of ongoing political debate. The basic introduction to the economics of healthcare should help you become a more informed participant in what will be continuously discussed at the national level for many years to come.

Objectives

At the end of this unit, you should be able:

- ✓ to know the definition of health economics;
- ✓ to understand the main characteristics of the healthcare market;
- ✓ to understand various problems in the healthcare sector.

Main Content

1.1. The Discipline of Health Economics

Health economics as an independent scientific discipline started more than seven decades ago in the sense that a specific consideration of topics related to the economics of the healthcare sector has become generally accepted. Currently, this field has proven itself so well that it has appeared in the regular curriculum of most

universities, and academic health economists work at medical departments, ties with economics itself are being strengthened, and the methodology applied is being improved. That there is a need for concern about the economic situation of the healthcare sector no longer requires long explanation. One of the main topics of health economics – and, therefore, the one that we will deal with most of all – is to find out what is obtained in terms of the outcome from the quite significant healthcare costs, and – in a slightly more sophisticated version – to develop methods which ensure as much as possible the maximum result using the given tools. Health economics has a broader scope than the study of a particular sector of the economy; the healthcare sector is not just another sector (like agriculture, industry or, say, financial services); its results are somewhat elusive, but they certainly go beyond what can be reasonably measured in monetary terms, due to the fact that the end result is individual health, or, to be more specific, an improvement in individual health. These are quantities that are difficult to compare between individuals and cannot be measured in monetary terms.

This special nature of the sector gives rise to many fundamental problems, which by themselves represent challenges to the economic theory. Health economics is by no means trivial. Many applications of health economics are of the type, in which deeper theoretical considerations are not widely used (although perhaps they should be used); indeed, much of the current interest in health economics has grown out of specific needs to be satisfied here and now (such as the advantage or disadvantage of using a new treatment, which may be more expensive than previous treatments, but may reduce other types of cost). The day-to-day nature of such considerations implies that too much theorizing should be avoided. On the other hand, much of the methodology applied in such cases depends on conventions, which ideally should have a theoretical foundation, and the need for such a foundation will surface from one time to another. One of the examples of this phenomenon is the controversy between supporters of human-capital versus frictional methods in assessing the "production gain" of a treatment: If the treatment makes the patients able to work more, how should this be measured, by wages earned or by the amount saved from

not having to call in an unemployed and giving the necessary instructions (the method chosen makes a big difference in the result). At a closer look, it turns out that health economics cannot easily be defined; as we argued above, it is not just the economics of the established healthcare sector, which, by the way, is not a very well defined concept since healthcare institutions, providers, and financing differ among countries. Also, an attempt to define the field by its output, that is "health", seems to be largely unsuccessful, perhaps due to the ambiguities in the concept of "health", which in many contexts is interpreted in a very wide sense so as to become synonymous with "welfare" or "happiness".

A well-known definition of "health" proposed by the World Health Organization (WHO) characterizes it as a state of perfect physical, moral and social well-being. It is questionable whether such a view of health will be productive when analyzing how much health we get from the healthcare costs. Indeed, it seems that using this definition, there is no longer a specific field of health economics since it has become synonymous with economics as such. Fortunately, there is no need for a precise definition of health economics since our primary interest will be in specific topics whether they happen to be typical of health economics or not. In that case, even "health" is the term, which we do not need very much. Rather, we will repeatedly consider models, in which we try to cover one of the supposedly very many different aspects of health. Our lack of enthusiasm in measuring health or even discussing the concept of "health" is, of course, due to the fact that we are engaged in economics, trying to be as accurate as possible with regard to the concepts included in the model, and also trying, as far as possible, to construct models in such a way that only those concepts are used, which are of immediate sense – which "health" uses only very rarely – enter as variables to be studied. To understand the role of economics in relation to healthcare, we must return to the basic structure of economic science and its functions. Economics deals with the description of the relationships between different individuals and organizations related to the production and consumption of goods and services. The main point of the study of these relationships is to explain how the institutional framework, the rules of conduct specified for individuals and organizations affect the final outcome.

Classical economic disciplines as the price theory and the welfare theory investigate the market mechanism; industrial organization focuses on the effects of imperfect competition on prices, welfare, and incomes. The theory of international trade studies the effect of different rules for international commodity exchanges, etc. At this level, it should not be surprising that health economics can be viewed as an economic discipline that deals with the institutional framework of healthcare (consumption, provision, financing) and the relationship between rules and institutions on the one hand, and the resulting health status of the population on the other hand. This still remains a somewhat vague description of the area and it seems difficult to approach it in a few words.

The **problems in the healthcare sector** will allow you to understand a lot about the economic theory. Below are some of the relevant areas.

1. Consumer substitution is one of the topics taught in economics – commodities compete with each other for the consumer's budget, and changes in the initial conditions (prices, budget, and tastes) will produce responses in the demand for all commodities. Substitution is a fundamental phenomenon in economics, in the medical profession, the viewpoint that health should be absolutely evenly distributed in the population is very firmly rooted. Although there seems to be no similar quest for equality in incomes, the fact that these two are interrelated comes as a big surprise. A striking example of substitution with unexpected health effects may be provided by an investigation of teenager behavior with respect to the use of mobile phones and smoking: while the use of mobile phones has increased dramatically, smoking habits have changed so that there are fewer smokers. A possible explanation is that both types of consumption have the main goal of signaling adulthood, but once teenagers engage in buying a mobile phone and using it, the budget no longer allows smoking, which is consequently reduced. The classical model of long-term consumption and individual health behavior by Grossman (1972a) is a story about the substitution. You can invest in your own health (by choosing the right diet, workout and frequent visits to the gym), and this investment will give you a payoff in terms of less time wasted on treating and curing your illnesses, but you will have to compare with other investments,

such as buying shares, which may or may not give you a better payoff.

2. In the healthcare sector, both consumption and production is subject to externalities. It matters to us what other people do or perhaps do not do. First of all, there is a simple externality connected with infectious diseases where the treatment of any patient has an effect on the number of possible future cases, thus, on the probability of any other person getting the disease. But the consumption externalities go beyond this. We experience disutility from seeing that other people do not get the same treatment for illness as we do ourselves, which means that our satisfaction depends on the consumption of other people besides ourselves. This is not in itself outstanding; traffic economists deal with congestion effects: the fact that so many people use their car has a detrimental effect on the pleasure that others get out of using their car. Also, the utility of conspicuous consumption (derived from showing other people that you can afford goods which they cannot) is reduced the more people engage in it. But in the healthcare sector the externality is other way round, and it is a factor to be considered in the design of a system of healthcare financing.

3. On the production side of the economy there is an element of natural monopolies – hospitals need a certain minimal size to function, and the cost structure is characterized by the presence of large fixed costs. There are other types of monopolies that are perhaps less based on technological characteristics and more on tradition and political expediency. Pharmacies have a monopoly on the sale of prescribed drugs, the medical industry produces under a monopoly based on patent rights. It is easily seen that market failure must be a central theme in any discussion of the economic performance of the healthcare sector.

4. Uncertainty is an important aspect of almost all economic behavior, but in some situations (actually, most of the situations treated by economic theory), it is acceptable to disregard it when studying the basic patterns of behavior. However, when dealing with problems of illness and treatment for illness, uncertainty is central to the problem. Consumption of this type of healthcare is consumption under uncertainty, and as such it must be considered in the proper perspective. It has been argued that the presence of

user payments for treatment does not reduce demand once the need for treatment is established – a broken leg must be treated whether the treatment is cheap or not. However, this argument neglects that consumption under uncertainty should be considered as a contingent consumption (depending on whether an illness occurs or not) and that there is a wide spectrum of choice available to the individual in determining the proper contingent contract (insurance). The notion of user payments cannot be understood separately from insurance and the types of market failure pertaining to insurance contracts related to asymmetric information in one of its several forms.

We can go on with this, showing that the diverse fields of the economic theory come into play in health economics, but it is better to proceed directly to health economics proper, where we shall consider the details with these and many other problems. The goal is not only to identify the problems and their theoretical content, but also to relate to the field of regulation and control. This is in many cases quite clear since markets for healthcare often do not regulate themselves; there is a need for regulation in the interest of society. Indeed, the healthcare sectors are highly regulated in most countries. Control and regulation are a central aspect of the economic organization of the healthcare sector. When the varying degree of direct public engagement in healthcare provision is added, it becomes clear that it is something that matters much. We will consider the merits and demerits of government engagement versus decentralized market solutions, and since our discussion will have another point of departure (namely the economic theory) than its counterpart in the public debate, the conclusions may not always be the same.

1.2. The Special Characteristics of the Market for Healthcare

The standard theory of how markets work is the model of supply and demand. This model has several notable features:

1. The main interested parties are buyers and sellers at the market.
2. Buyers are good judges of what they get from sellers.
3. Buyers pay sellers directly for the goods and services being exchanged.

4. Market prices are the primary mechanism for coordinating the decisions of market participants.

5. The invisible hand, left to its own devices, leads to an efficient allocation of resources.

For many goods and services in the economy, this model offers a good description. But none of these features of the standard model reflects what goes on in the healthcare market. Like other markets, the healthcare market has consumers (patients) and producers (doctors, nurses, etc.). But various features of this market complicate the analysis of their interactions, in particular:

– Third parties – insurers, governments, and unwitting bystanders – often have interesting healthcare outcomes.

– Patients do not know what they need and cannot evaluate the treatment they are getting.

– Healthcare providers are often paid not by the patients, but by private or government health insurance.

– The rules established by these insurers, more than market prices, determine the allocation of resources.

– In the light of the above four points, the invisible hand cannot work its magic, and therefore, the allocation of resources in the healthcare market may not be effective.

Healthcare is not the only product or service in the economy that departs from the standard model of supply, demand, and the invisible hand. But healthcare may be the most important product or service that departs so radically from this benchmark. Examining the special features of this market is a good starting point for understanding why the government plays a great role in the provision of healthcare and why health policy is often complex.

(a) The Prevalence of Externalities. Market outcomes may be inefficient when there are externalities. An externality arises when a person engages in an activity that affects the well-being of a bystander, but neither pays nor receives compensation for that effect. If the impact on the bystander is adverse, it is called a negative externality. If it is beneficial, it is called a positive externality. In the presence of externalities, the society's interest in a market outcome extends beyond the well-being of buyers and sellers who participate in the market to include the well-being of bystanders who are affected indirectly. Since buyers and sellers neglect the external effects

of their actions when deciding how much to demand or supply, the externality can render the unregulated market outcome inefficient. This general conclusion is crucial for understanding healthcare because externalities in the market are so common. These externalities can call for government action to eliminate the market failure. Consider vaccines, for example, if one person vaccinates himself/herself against a disease, he/she is less likely to catch it. But because he/she is less likely to catch it, he/she is less likely to become a carrier and infect other people. Thus, getting vaccinated conveys a positive externality. If getting vaccinated has some cost, either in money, time, or risk of adverse side effects, too few people will choose to get themselves vaccinated because they will likely ignore the positive externalities when weighing the costs and benefits. The government may remedy this problem by subsidizing the development, manufacture, and distribution of vaccines or by requiring vaccination. Another example of an externality in the healthcare system concerns medical research. When a physician figures out a new way to treat an ailment, that information enters the society's pool of medical knowledge. The benefit to other physicians and patients is a positive externality. Without the government intervention, there will be too little research. Governments respond to this externality in many ways. Sometimes, the government grants the researcher a patent on the new product, as is the case with new pharmaceutical drugs. The patent gives an incentive for research because the patent holder can profit from a temporary monopoly. The patent is said to internalize the externality. Yet this approach is not perfect as the monopoly price is higher than the marginal cost of production. The high monopoly price reduces the consumption of the patented treatment, leading to inefficiency as measured by the dead weight loss. Moreover, the high price may be hard on lower-income patients. Another approach to dealing with the positive externality from medical research is for the government to subsidize the research. This policy requires taxation to raise the necessary funds, and taxation entails deadweight losses of its own. But if the externalities from the funded research exceed the cost of the research, including the deadweight losses, overall welfare will increase.

(b) The Difficulty of Monitoring Quality. In most markets, consumers know what they want, and after a transaction is completed,

they can judge whether they are happy with what they got. Healthcare is different. When you get sick, you may not know what the best treatment is. You rely on the advice of a physician, who has years of specialized training. And even with hindsight, you cannot judge for yourself whether the treatment the physician offered you was the right one. Sometimes state-of-the-art medicine fails to improve the patient's health. And given the natural restorative power of the human body, the wrong treatment can sometimes appear to work. The inability of healthcare consumers to monitor the quality of the product they are buying leads to various regulations. Most importantly, the government requires physicians, dentists, nurses, and other health professionals to have licenses to practice. These licenses are granted only after an individual attends an approved school and passes rigorous tests. Similarly, the Food and Drug Administration (e.g. NAFDAC) oversees the testing and release of new pharmaceutical drugs to make sure they are safe and effective. In addition to the government regulation, the medical profession monitors itself by accrediting medical schools, promoting best practices, and establishing professional norms of behavior. A physician's advice is supposed to be based on the patient's best interest, not on the physician's personal gain. When patients accept the advice, they rely on a degree of trust, which is often fostered by long-term relationships between a doctor and a patient. Suspicions about the standard economic motive of self-interest and the role of trust in healthcare relationships may explain the prevalence of non-profit hospitals. In some ways, hospitals are like hotels, but while most hotels are for-profit businesses, most hospitals are run by the government or established as non-profit entities. When consumers cannot judge the quality of the product they are buying, they may be more willing to trust an institution that is not set up primarily to enrich its owners. These public and private regulations of healthcare have their critics. For example, some economists have argued that there are too many hurdles to opening new medical schools. They suggest that the medical profession acts like a monopoly. By restricting the number of doctors, it drives up doctors' salaries and consumers' healthcare costs. Other economists have argued that the food and drug administration is too slow in approving new drugs.

Some patients who might have benefited from experimental treatments are forced to go without them.

1. The Insurance Market and Its Imperfections. Spending on healthcare is unpredictable because people do not know when they are going to get sick or what kind of medical treatments they will need. This uncertainty, and how people respond to it, is a key reason why we have the health institutions. Some of the issues with health insurance include the Value of Insurance. Most people are risk averse, that is, they dislike uncertainty. Imagine that you are faced with a choice between a certain income of N$100, 000 and a 50-50 probability of income of N$50,000 or N$150, 000. The two options offer the same average income, but the second is riskier. If you prefer the certain N$100, 000, you are risk averse. The same behavior arises from the randomness of health spending. Suppose that some disease affects 2 percent of the population and that everyone is equally likely to be afflicted. Treatment costs N$30,000. In this case, the expected cost of healthcare is 2 percent of N$30,000, which is N$600. If people are risk averse, they prefer to pay N$600 with certainty over a 2 percent chance of having to pay N$30,000. Giving people this option is the purpose of insurance. The general feature of insurance contracts is that a person facing a risk pays a fee (called a premium) to an insurance company, which in return agrees to accept all or part of the risk. Health insurance covers the risk of expensive medical treatment. In our example, a health insurance company can charge a premium of N$600 (or slightly more to make a profit) in exchange for promising to cover the cost of the N$30,000 treatments for 2 percent of its customers who get the disease. Markets for insurance are useful in reducing risk, but two problems impede their ability to do so efficiently.

2. Moral Hazard. The first problem that hinders the operation of insurance markets is moral hazard. When people have insurance to cover their spending on healthcare, they have less incentive to engage in behavior that will keep that spending to a reasonable level. For example, if patients do not have to pay for each visit to a doctor, they may go too quickly when they experience minor symptoms (a runny nose, an achy finger). Similarly, physicians may be more likely to order tests of dubious value when they know an

insurance company is picking up the bill. Health insurance companies try to reduce the problem of moral hazard by finding ways to encourage people to act more responsibly. For instance, rather than picking up the entire cost of a visit to a physician, they may charge patients co-pays of, say, N$20 per visit to deter patients from making unnecessary visits. Similarly, insurance companies may have strict rules about the circumstances under which they will cover the cost of certain tests that physicians order.

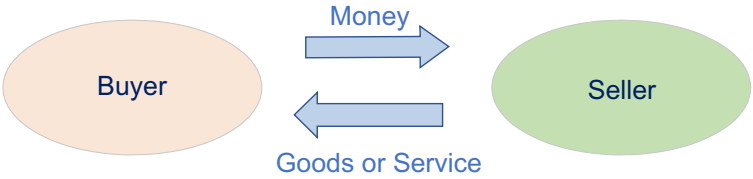
3. Adverse Selection. The second problem that impedes the operation of insurance markets is adverse selection. If customers differ in their relevant attributes (such as whether they have a chronic disease) and those differences are known to customers, but not observable by insurers, the mix of people who choose to buy insurance may be expensive to insure. In particular, people with greater hidden health problems are more likely to buy health insurance than are healthy people. As a result, for an insurance company to cover its costs, the price of health insurance must reflect the cost of a sicker-than-average person. Even people with average health may see the high price and decide to go without insurance. As people drop coverage, the insurance market fails to achieve its purpose of eliminating the financial risk from illness. Adverse selection can lead to a phenomenon called the death spiral. Suppose that since a person's health profile is private information, insurance companies must charge everyone the same price. At first, it might seem to make sense for a company to base the price of insurance on the health characteristics of the average person in the population. But after it does so, the healthiest people may decide that insurance is not worth the cost and drop out of the insured pool. With a sicker group of customers than expected, the company has higher costs and, therefore, has to raise the price of insurance. The higher price now induces the next healthiest group of people to drop insurance coverage, which drives up the cost and price again. As this process continues, more people drop coverage, the insured pool gets less healthy, and the price keeps rising. In the end, the insurance market may disappear. The problem of adverse selection has been central in the debate over health policy.

(c) Healthcare as a Right. Normally, when some people do not buy a product or service, perhaps, as they think it costs too much

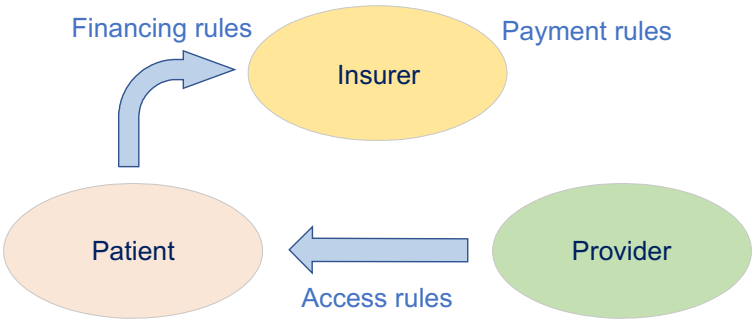
taking into account their income, such a result is not a major problem for society. For example, suppose that a ticket to a cinema becomes expensive and lower-income consumers choose other forms of entertainment. We may argue that good theatre is not enjoyed more widely, but few would argue that this is a great injustice. Healthcare is different. When a person gets sick, it seems wrong that a low income would be a reason to deny treatment. Healthcare, unlike a ticket to the cinema, is perhaps a human right. This judgment goes beyond the scope of economics and is best left to political philosophers, but we should acknowledge this belief as we study the economics of healthcare. In some ways, healthcare is like food. Food is necessary to survive, and as a society, we try to ensure that everyone has the resources to get the food they need. There is, however, an important distinction between food and healthcare. Over time, the price of food has risen more slowly than incomes, and so affording an adequate diet has taken up a declining share of the typical household's budget. On the contrary, as the cost of state-of-the-art healthcare is growing rapidly, its provision requires an increasing share of the budget of a typical family. Healthcare being viewed as a right, along with its rising cost, has led to a large role for the government. In many countries, the government runs the healthcare system, financed mostly by taxes. This system is sometimes called single-payer because one entity – the government's health service – pays all the bills. By contrast, in the United States and Nigeria, most people have private health insurance, often through their employers, but the government still has a sizable presence. In the United States, Medicare provides health insurance for those 65 and older; Medicaid provides health insurance for the poor; and the Affordable Care Act regulates the market for private health insurance and gives insurance subsidies to many lower-income households. There is a little doubt that, with healthcare often viewed as a human right, the government will continue to play a large role in the healthcare system.

(d)The Rules Governing the Healthcare Market Place: The importance of health insurance, whether provided by private companies or the government, requires that the market for healthcare works differently than most other markets in the economy. The typical markets – say, the market for rice – looks like panel (a) of Fig. 1.1.

The market has buyers and sellers. A seller offers a product or service at a price. A buyer who wants the item simply has to offer the right amount of money. An exchange is made, and soon the seller is counting the profit, and the buyer enjoys his product.



a) Model of the typical market



b) Model of a Healthcare market with an Insurer

Fig. 1.1. Models of Typical Market and Healthcare Market

The market for healthcare looks more like panel (b). The provider (the seller of medical services) is not paid directly by the patient (the buyer). Instead, the patient pays money to an insurer in the form of either a premium (if the insurer is a private company) or taxes (if the insurer is the government). The insurer uses this money to compensate the provider, who, in turn, provides medical services to the patient. This process requires three sets of rules to guide behavior. The first set determines the financing, that is, who pays for the insurance and how much they pay. If the insurer is the government, paying for healthcare becomes part of designing the tax system. If the insurer is a private company, healthcare is financed by the price that health insurance purchasers pay for

their coverage. The price is set at the insurance market, which (like other markets) bases price on costs. In many cases, however, state and federal governments regulate the market for private insurance. For example, they may limit the extent to which companies can charge different prices based on age, sex, and pre-existing conditions. Thus, even when the financing of healthcare occurs between a patient and a private insurer, it is still shaped by public policy. The second set of rules determines a patient's access to healthcare. Since insured patients do not pay the marginal cost of each medical service they consume, there is the possibility of overuse (moral hazard). To mitigate this problem of moral hazard, the insurer (whether the government or a private firm) has rules to limit access to when it makes sense. In other words, these rules ration the use of medical services based on estimated costs and benefits. For example, a patient may be able to get a routine check-up no more than once a year, may have access to only a limited number of doctors, or may need a referral from a general practitioner before making an appointment with a more expensive specialist. Such access rules are necessary since once people have insurance to pick up the cost, market prices are no longer giving them the right signals about how to allocate scarce resources. The third set of rules determines the payments from insurers to providers. These rules establish both what an insurer will pay for and how much he will pay. Treatment prices affect which treatments providers refer patients to. Insurers may deem some treatments too expensive, too experimental, or insufficiently valuable to pay for them at all. In such cases, providers will often not offer patients the services. Sometimes, however, providers will offer the services only if the patient pays the full cost of the treatment (as is often true with cosmetic procedures). In this case, the market for healthcare reverts from panel (b) in Fig. 1.1 to the more typical market in panel (a). The rules regarding financing, access, and payment are related, and together they shape the kind of healthcare system a nation has. For nations with a government-run system, these rules are set by public policy. For nations with more private insurance, such as the United States, these rules are set by insurance companies as they compete for customers, subject to various government regulations.

1.3. Measuring Health

Can health status be measured? Intuitively, it is clear that a closer analysis of the use of resources for improving health conditions, for society or single individuals, will depend on how the state of health is measured. It would be very helpful if a numerical measure of health were available, so that a "marginal health effect" of each conceivable treatment might be computed as change in health per monetary value spent in the treatment. As already mentioned, there are considerable difficulties connected with such a measurement. There is no obvious unit of measurement for health, and even the concept of "health" as such is not clear. This should not be a cause of despair since most of the economic disciplines run into similar difficulties. Even when seemingly exact measures exist, problems show up at a closer analysis (for example, in national accounts: what does GNP actually measure?). On the other hand, it is clear that the analysis improves with more precise measures of the consequences of economic choices. Therefore, it is important to investigate how far one can get in measuring health. This measurement problem encompasses all of health economics. At the outset it is easily seen that there can be no measurement of health corresponding to those of the national accounts (where it makes sense to consider differences of two measured values as an expression of the magnitude of the improvement), but one might still hope for constructing a suitable scale and positioning different health states on this scale in such a way that higher scale value corresponds to better health. There is also a problem of interpersonal comparisons – is it possible to compare the measures of health of two persons, concluding that one of them has a better state of health than the other – and further on, can we aggregate the health of a whole society and then compare the health state of two different countries? It may be seen from this discussion that it creates a more detailed argument about the nature of the scales, on which health is to be measured (a discussion known from the distinction between cardinal and ordinal utility in the consumer theory). Therefore, an overview of the methods for measuring health employed in practice

is considered. Since health *a priori* is something ranging from perfectness to the total absence of health (death), a scale for measuring health states can naturally be chosen as the interval of real numbers from 0 to 1. The approach taken is as follows. First of all, some fundamental characteristics of health are highlighted; each of them describes certain aspects of health. The degree of fulfillment of the demand for perfect health in each of these aspects is then measured on a scale from 0 to 1 (or from 0 to 100). The difficult part of the measurement is then the weighing together of the scores in each of the health characteristics. To do this, a panel of individuals are questioned about the trade-offs between different states of health (where health is perfect in all except one of the aspects), and the average score is used to sum the scores for each of the aspects into an overall health assessment. A total of eleven characteristics were chosen: the ability to move around, the ability to hear, the ability to talk, sight, the ability to work, breathing, incontinency, the ability to sleep, the ability to eat, intellectual and mental functioning and the social activity. For each of these characteristics a numerical value is determined belonging to a precisely described state of imperfect functioning.

UNIT 2

THE FOUR BASIC QUESTIONS IN ECONOMICS. PUBLIC HEALTH AND HUMAN NEEDS IN SOCIETY

Content

Introduction

The study of health economics involves the application of various microeconomics tools, such as demand or cost theory to health issues and problems. The goal is to promote a better understanding of the economic aspects of healthcare problems so that corrective health policies can be designed and proposed. A thorough understanding of the microeconomic analysis is essential for conducting a sound analysis of health economics. The tools of health economics can be applied to a wide range of issues and problems pertaining to health and healthcare. For example, the analysis of health economics might be used to study why 29 out of every 1,000 babies born in Nigeria died before their first birthday, whereas all but 3 out of 1,000 babies born in Japan live to enjoy their first birthday cake. The tools of health economics can also be used to understand the economic desirability of a contested merger between two large

hospitals in a major metropolitan area. The main question is if the merger of the two hospitals results in lower hospital prices due to overall cost savings or higher prices due to monopoly power.

Health economics can be seen as the application of economic theories, tools and concepts of economics as a discipline to the topics of health and healthcare. Since health economics is concerned with issues relating to the allocation of scarce resources to improve health, this includes both resource allocation within the economy to the health sector and within the healthcare system to different activities and individuals.

Health economics is difficult to define in a few words since it encompasses such a broad range of concepts, theories, and topics. The Mosby Medical Encyclopedia defines health economics as follows: "**Health economics** *studies the supply and demand of healthcare resources and the impact of healthcare resources on the population.*"

Note that health economics is defined in terms of the determination and allocation of healthcare resources. This is logical because medical products cannot exist without them. Healthcare resources consist of medical supplies, such as pharmaceutical products, personnel, such as physicians and lab assistants, and capital inputs, including nursing homes and hospital facilities, diagnostic and therapeutic equipment and other items that provide medical care services. Unfortunately, healthcare resources, like resources in general, are limited or scarce at a given time, and wants are limitless. Thus, trade-off is inevitable, and a society whether it possesses a market-driven or a government-run healthcare system must make a number of fundamental and crucial choices. These choices are usually based on four basic questions:

1. What combinations of non-medical and medical products and services should be produced in the macroeconomy?

2. What particular medical products and services should be produced in the health economy?

3. What specific healthcare resources should be used to produce the chosen medical products and services?

4. Who should receive medical products and services that are produced?

Objectives

At the end of this unit, students should be able to understand:

✓ allocative and production efficiency in healthcare;

✓ distributive efficiency in healthcare;

✓ implications of the ways that society chooses to answer allocative and distributive questions;

✓ health as one of the social sectors with economic implications;

✓ the specific nature of the healthcare service in implementing economic principles and techniques.

Main Content

2.1. Implications of the Four Basic Questions

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

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

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

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

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

2.2. Production, Allocative Efficiency and Production Possibility Curve

The most straightforward way to illustrate production and allocative efficiency is to use the production possibility curve (PPC) – an economic model that depicts the various combinations of any two goods or services that can be produced efficiently given the stock of resources, technology, and various institutional arrangements (Fig. 2.1)

The PPC shows the trade-off between any two products given a fixed stock of resources and technology. Any point on the PPC, such as points A through E, reflects efficiency since units of one product must be given up to receive more of the other. A point in the interior, such as F, reflects inefficiency as more of one product can be attained without necessarily reducing the other. A point outside the PPC, such as G, is not yet attainable, but can be reached with an increase in resources or through institutional or technological changes that improve productivity.

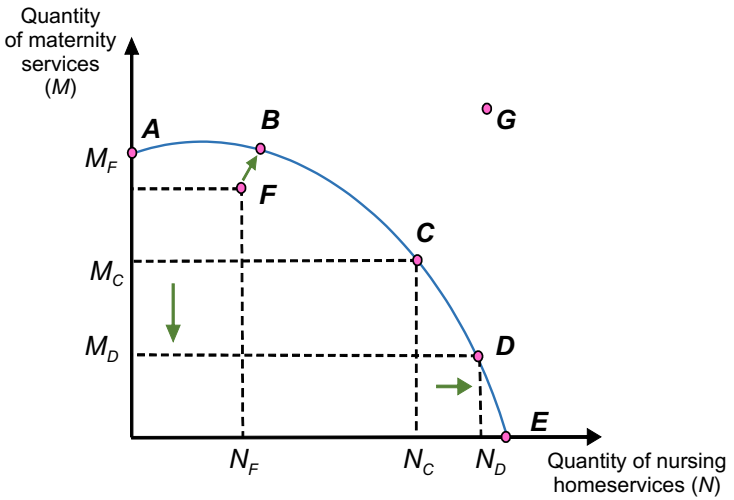


Fig. 2.1. Production Possibility Curve for Maternity and Nursing Home Services

The quantities of maternity services, M, and nursing home services, N, are shown on the vertical and horizontal axes, respectively. Three points on the bowed-out PPC depict the various combinations of maternity and nursing home care services that can be efficiently produced within the health economy assuming the amounts of healthcare resources and technology are fixed at a given point in time.

Every point on the PPC implies production efficiency since all healthcare resources are being fully utilized. For example, note points A, B, C, D, and E on the PPC. At each of these points, medical inputs are neither unemployed nor underemployed (for example, a nurse involuntarily working part-time rather than full-time) and are being used in the most productive manner so that society is getting their maximum use. If a movement along the curve from one point to another occurs, units of one medical service must be forgone to receive more units of the other medical service. Specifically, assume the health economy is initially operating at point C with MC units of maternity care services and NC units of nursing home services. Now suppose healthcare decision-makers decide that society is better off at point D with one more unit of nursing

home services, ND – NC. The movement from point C to point D implies that MC – MD units of maternity care services are given up to receive the additional unit of nursing home services. Since medical resources are fully utilized at point C, a movement to point D means that medical inputs must be drawn or reallocated from the maternity care services market to the nursing home services market. As a result, the quantity of maternity care services must decline if an additional unit of nursing home services is produced. The forgone units of maternity care services, MC – MD, represent the opportunity cost of producing an additional unit of nursing home services. Generally, opportunity cost is the value of the next best alternative that is given up. The bowed-out shape of the PPC implies that opportunity cost is not constant, but increases with a movement along the curve. Imperfect substitutability of resources is one reason for this so-called law of increasing opportunity cost. For example, suppose the nursing home services market expands downward along the PPC. To produce more nursing home services, employers must bid resources away from the maternity care services market. Initially, the least productive inputs in the maternity care services market are likely to be bid away as they are available at a lower cost to nursing home employers. Consequently, very few maternity care services are given up at first. As the nursing home services market continues to expand, however, increasingly productive inputs in the maternity care services market must be drawn away. The implication is that society gives up ever-increasing units of maternity care services. Thus, the law of increasing opportunity cost suggests that ever-increasing amounts of one product must be given up to receive successively more equal increments of another product. If medical inputs are not fully utilized since some inputs are idle or used unproductively, more units of one medical service can be produced without decreasing the amount of the other medical service. An example of an underutilization of resources is indicated by point F in the interior of the PPC. At point F, the healthcare system produces only MF units of maternity services and NF units of nursing home services. Note that by moving to point B on the PPC, both maternity care services and nursing home services can be increased without decreasing the other. The quantities of both products increase only because some resources are initially

idle or underutilized at point F. Healthcare resources are inefficiently employed at point F. A point outside the current PPC, such as G, is attainable in the future if the stock of healthcare resources increases; a new, productivity-enhancing technology is discovered; or various economic, political, or legal arrangements change and improve productive relationships in the health economy. If so, the PPC shifts out and passes through a point like G. For example, the technological change may enable an increased production of both maternity and nursing home services from the same original stock of healthcare resources. Alternatively, a greater quantity of maternity and nursing home services can be produced, and the PPC shifts outward if more people enter medical professions (possibly at the expense of all other goods and services). Production efficiency is attained when the health economy operates at any point on the PPC since medical inputs are producing the maximum amount of medical services and no unproductive behavior or involuntary unemployment exists. Allocative efficiency is attained when society chooses the best or most preferred point on the PPC. All points on the PPC are possible candidates for allocative efficiency. The ideal, or optimal, point for allocative efficiency depends on the society's underlying preferences for the two medical services.

2.3. Public Health and Human Needs in Society

Humans are all very diverse. We come from various ethnic groups, countries, backgrounds, genders, and belief systems. We have different types of interests, careers, hobbies, and lifestyles. However, as diverse as we are, we all share the same basic human needs. The original concept of "basic human needs" was written about by a humanistic psychologist Abraham Maslow (Fig. 2.1). He wrote about what he called the "hierarchy of needs" in his Motivation and Personality.

The idea was to study the psychology of healthy-minded people. Here is how Maslow breaks down the basic human needs.

1. **Survival: Physiological Needs.** The most basic human need is to survive. We need food, water, air, clothing, shelter, and sleep. If these needs are not met, we are not going to survive for very long.

2. **Safety: Safety Needs.** Next, we need to feel safe and secure. We want to be physically safe from violence and danger. We seek job security and financial stability. Emotionally, we want stable, loving relationships with our family, friends, and coworkers.

3. **Society: Love Needs.** We need a connection with others. A sense of belonging. We congregate. We join groups and clubs, seeking out others with similar interests. This need both to give and receive affection is so overpowering that it can even override our need for safety, causing us to stay in abusive relationships (see also: Stockholm Syndrome).

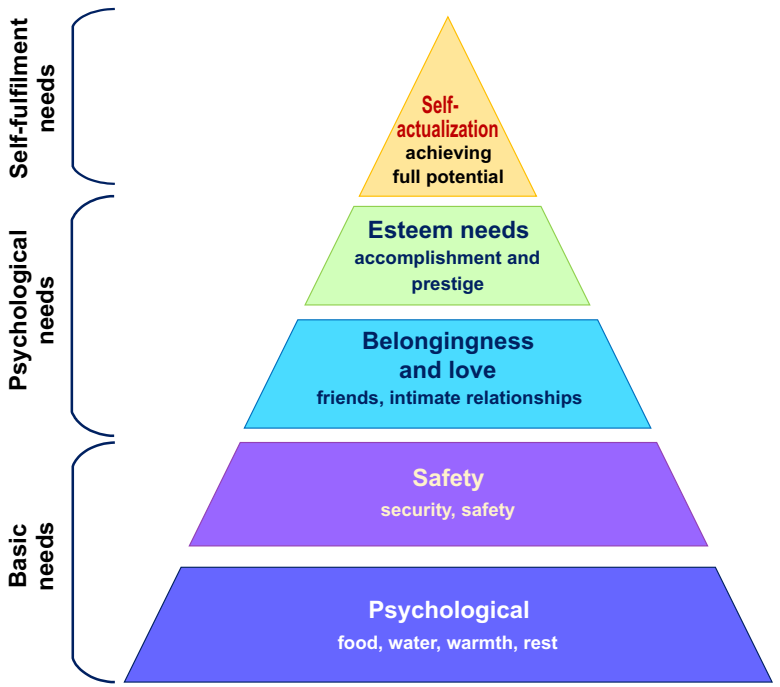


Fig. 2.2. The Maslow's Pyramid of Human Needs in Society

4. **Swagger: Esteem Needs.** The fourth need works at two levels, what Maslow called a lower esteem (the need for others to respect us) and a higher esteem (the need for self-respect). It is the need for self-confidence, independence, and the right to self-determination. We want to know that our contributions are valued and meaningful. We want to make a difference in the world.

5. **Self-Actualization: the Need to Grow.** Maslow first called this self-actualization, then later self-transcendency. Once we have the basic needs covered, we seek to master those needs. We set our sights higher, and work to become the very best self that we can be. We are not content to sit back and be the same person this year that we were last year. That is why New Year's resolutions are so popular: we are driven to improve ourselves.

A health economy, like a macroeconomy, involves the production and consumption of goods and services and the distribution of those goods to consumers. A health economy differs from a macroeconomy since it distinctly considers production, consumption, and distribution activities that directly relate to the population health. Another difference concerns the way, in which economists take the pulse of the macroeconomy and health economy. While economists are really concerned with efficiency and employment equity, inflation, and gross domestic product growth rates are also considered when evaluating the effectiveness of a macroeconomy. If you recall from ECON 100, the gross domestic product (GDP) captures the total market value of all goods and services produced in an economy during a particular period. For a health economy, the analogous performance indicators are the components that make up the so-called three-legged stool of medical care: costs, access, and quality.

As mentioned earlier, health, like any other durable product, generates a flow of services. These services yield satisfaction, or what economists call utility. Your television set is another example of a durable product that generates a flow of services. It is the many hours of programming, or viewing services, your television provides that yield utility, not the set itself. As a good, health is desired for consumption and investment purposes. From a consumption perspective, an individual desires to remain healthy because she or he receives utility from an overall improvement in quality of life. In simple terms, a healthy person feels great, and thus, is in a better position to enjoy life. The investment element concerns the relation between health and time. If you are in a positive state of health, you allocate less time to sickness, and therefore, have more healthy days available in the future to work and enhance your income or to pursue other activities, such as leisure.

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

2.4. General Features of Healthcare

There are different understandings of health – each with different implications for the roles of the government. It is important to recognize, first, the difference between health and healthcare. The term "health" refers to a state either of an individual or of a community. This state of health can be influenced by a number of factors, including "healthcare". However, other factors that affect health are poverty, level of education, food intake, access to clean water and sanitary and housing conditions. The narrowest concept of health sees it as a measure of the state of the physical body organs. An individual is unhealthy if there is a malfunctioning of part of the body. A broader, but related, definition sees health just in terms of the mechanics of the different bodily organs, but in the ability of the body as a whole to function.

In contrast, the WHO definition of health as "a state of physical, mental and social well-being and not merely the absence of disease or infirmity" indicates a clear shift away from earlier narrow organic or functionally-based definition of health to a more holistic view, it sees the health of an individual or community as being concerned not only with physical (and mental) status, but also with social and economic relationships.

Individual conceptualization of health will affect the type of intervention and planning that is possible. The narrowest definitions are closely associated with a medical model of health, in which the role of health services is seen as paramount in restoring the functioning of the unhealthy body. Wider primary healthcare concepts suggest

that broader interventions, including community empowerment and anti-poverty measures, are necessary to promote health. Three perspectives can be used to distinguish health on the importance of health and on the possible roles of the state in promoting it.

(a) *Health as a right*. Some see this as a right similar to justice or political freedom. Indeed, the WHO constitution states that "… the enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition". Although it is difficult to believe that the equal health status is attainable in the same way that the equal political freedom may be, health is seen as so fundamental that constraints to its full attainment must be minimized. In part, this involves ensuring access to healthcare. The government is seen as having a responsibility to ensure this, comparable with its role in ensuring equal justice. According to such a view, a government will be particularly concerned with issues of equity in health and healthcare.

(b) *Health as consumable good*. Health is also seen as an important individual objective that is not comparable with justice, but rather with material aspects of life. Such a view often refers to health as a consumption product. The government here has no special responsibilities in the promotion of health, but leaves decisions as to its comparative importance to individual consumers. The role of the state under such a view might be limited to ensuring that the healthcare provided is of adequate quality (such as ensuring professional standards in the same way that it would monitor the quality of any good or service, such as food).

(c) *Health as an investment*. The third view of health is that it is important, but largely it affects the productive ability of the workforce. Illness may affect overall production, either through absenteeism or by lowering productivity through its debilitating effects.

Important! Distinctive Characteristics of Healthcare Services from other Commodities

Why not leave healthcare to the market? Most people believe that you cannot buy and sell healthcare like other goods and services. They believe that healthcare is different. This is what is sometimes called a "common-sense" approach to the issue. Economists approach the same question differently. For many people, the word

market conjures up a picture of a town square with lots of small stallholders selling everything, from fruit and vegetables to meat and fish. For economists, the term has a much wider meaning:

– It is used to describe any process of exchange between buyers and sellers.

– Formally, a market could be defined as any set of arrangements that allows buyers and sellers to communicate and thus arrange exchange of goods, services or resources.

– A free market is where such exchange occurs without interference from the government.

– Information is a vital ingredient for any market. Both buyers and sellers need to have access to sufficient information allowing them to make rational decisions.

In theory, markets produce goods and services in the right quantities and at the lowest possible cost. This is why markets are so powerful. Nevertheless, in the real world, markets do not always work in the way theory predicts. It is possible for a free market to produce a Pareto inefficient result, i.e. the market fails. A Pareto inefficient situation can occur as a result of the following things:

(a) Imperfect information. We get goods at the lowest possible cost if the market is able to transmit all the information about benefits and costs between producers and consumers. If this information is less than perfect, then the market will fail. Think about buying a CD. You know what a CD is, and you will have a good idea of the kind of music on the disc. Therefore, you are able to relate your benefit to the price of the CD. If we look at the market for CDs, people will go on buying CDs until the extra satisfaction from the last CD is exactly equivalent to the price of the CD. However, healthcare is rather different from CDs. We face very acute information problems, which make rational purchasing decisions difficult. For instance, most people do not know the best way to treat a stomach ulcer, so they would find it difficult to buy such treatment. This analysis also assumes that the only people receiving benefit or satisfaction from the CDs are the people buying them. In other words, the price of a CD accurately conveys the level of satisfaction received. This ignores the possibility of externalities or "spillovers". Think about someone hearing your CD and enjoying it – they are also receiving satisfaction from the disc, but the market is unable

to provide any information about the benefits they are receiving unless they specifically share the cost of buying the CD. Whenever externalities occur, the market fails. Many economists believe that there are strong externality effects related to healthcare. For example, caring for a sick person can impose financial costs on that person's family.

(b) Perfect Competition. An efficient free market requires producers to be operating under conditions of perfect competition. This requires a stringent set of conditions – perfect information, many buyers and sellers, a uniform product and freedom of entry and exit – which ensure that firms are price takers, producing for the lowest possible cost in the long run and only earning normal profits. If producers do not operate in this way and, in particular, if they have a significant power to influence the price or the total quantity being produced, then the market will fail. Doctors and other suppliers of healthcare often have this power.

(c) Problems of Risk and Uncertainty. If we are going to buy healthcare in a free market, then we have to have enough money to pay for it. Nevertheless, healthcare is expensive, and we cannot predict when we are going to be ill. What makes this worse is that postponing buying healthcare is often risky. So, we face the problems of risk and uncertainty. The market response to this problem is to develop an insurance market to remove the uncertainty and risk from healthcare spending. We pay an agreed amount of money per year whether we need healthcare or not. Then, when we need care, the insurer pays the bills, however large they are. So, a free market in healthcare requires an effective healthcare insurance market. Unfortunately, the healthcare insurance market itself is often not efficient. Moral hazard and adverse selection both cause significant market failure. Unequal information, moral hazard and adverse selection explain why a free market in health insurance is unlikely to be efficient. However, healthcare markets face even more fundamental information problems.

(d) Consumers as Satisfaction Maximizers. Are consumers the rational satisfaction maximizers? The market theory assumes that consumers know what is best for themselves, that is, they can make choices which will maximize their total satisfaction. If this assumption is wrong, then markets will not automatically produce efficient

results. The satisfaction gained depends on the quantity and mix of goods and services chosen. The theory assumes that consumers get more satisfaction from more goods and services, but that the increase in satisfaction from consuming another unit – the marginal utility – diminishes as consumption rises. "By choosing a particular bundle of goods, people show that they prefer it to all others; thus, it is best for them. In addition, if all people are in their best position, then society – which is simply the aggregation of all people – is in its best position. Therefore, allowing people to choose in the marketplace results in the best of all possible economic worlds". Thomas Rice in the Economics of Health Reconsidered suggests a range of reasons why this view of consumer behavior could be mistaken. These include:

1. The idea that consumer utility depends on the bundle of goods and services consumed. If this were true then people in rich developed economies ought to be happier than people in poor economies. But, research by Easterlin (1974) argued that utility depended on relative consumption, so rich people were happier than poor people in all societies. This means that if you consume more, it may reduce my utility because I am relatively worse off now.

2. The traditional theory ignores the issue of how tastes are determined. Evidence from social psychology suggests that tastes are determined by people's past and present environments. So, for instance, if you are in a peer group, which smokes, then you are likely to develop a "taste" for smoking, which will remain, even after you have left the peer group. If this is true, then it is not clear that satisfying tastes will actually make people better off. In fact, "if one believes that tastes are determined in such a way, it becomes clear that a society might be better off pursuing some goods and services that are not demanded mostly by the public. This is because people may not know what alternatives are available that will make them better off". Are consumers rational? What do economists mean by the concept of rationality? In a narrow sense, they mean that people will behave consistently – so, if they prefer A to B and B to C, then they will prefer A to C. More widely, they mean that people will behave in a reasonable manner. If consumers are not rational in this sense, then they will not necessarily make decisions, which maximize their welfare. Social psychology suggests that people are

often not rational in this sense – instead they exhibit what is called cognitive dissonance. In other words, they simultaneously hold two ideas that are psychologically inconsistent and use various forms of self-justification and rationalization to overcome the tension.

(e) Imperfect Competition. The free-market models predict large numbers of buyers and sellers, all of whom have no power individually to influence the market price. However, a significant proportion of healthcare is delivered by hospitals, and these hospitals can often exercise monopoly power within the healthcare market in the local area.

(f) Externalities. The economist defines external effects as involving positive and negative results for others that are the consequences of one's own actions. Externalities or spillover effects provide another source of market failure. Again, the problem is related to information. This time the market price does not accurately contain all the information about the benefits and costs of the market transaction.

The Political Economy of Healthcare. *A normative statement was usually based on the efficiency criteria of welfare economics. This raised the issue of whether a Pareto-optimal design of a healthcare system might ever be achieved. This section raises the question of what determines the actual (rather than any desired) institutional structure of a healthcare system. This type of question is the topic of Political Economy, also known as Public Choice. With regard to health policy and regulation, the following agents can be distinguished.*

Citizens. In a direct democracy, citizens may challenge a law that has been passed by a popular referendum. They may also force the legislature to deal with an issue through a popular initiative. In a purely representative democracy, voters have a mere indirect influence by voting for candidates for political office or parties who promise to pursue a certain policy. Even in a dictatorship, citizens are not without influence because at least some of them must be won over to keep public administration and the economy functioning. The more closely the health policy adopted by a dictatorial government matches the preferences of the citizenry, the less costly it is for it to maintain its power.

1. Politicians: In a democracy, politicians need to obtain votes. Promising to organize the provision of healthcare services (or at least

the availability of health insurance) may be a selling proposition. Meanwhile, younger voters may think that these public programs place a heavy financial burden on them while benefitting mainly the elderly.

2. Executive member of government: Gaining or maintaining executive power may calls for a great deal of financial support, which comes from large companies engaged in the healthcare sector (insurers, pharmaceutical companies) or professional associations (of physicians, nurses and hospitals). In general, the "supply side" tends to prevail in health policy at the governmental level.

3. International organizations: The World Health Organization (WHO) has had considerable success in influencing the national health policy by emphasizing the risks posed by epidemics. Increasingly, decisions affecting the health policy and regulation are made by the World Trade Organization (WTO), and in particular the European Union (EU). In both instances, the fact that traded commodities may have an impact on health while some health products are tradable provides a justification for intervention.

The focus here is on the viewpoint of citizens and voters. Their interests are decisive in countries with good governance since the other levels must take them into account in order to ensure their political survival

2.5. Health and Economic Development

The modern view of development perceives it as both a physical reality and the state of mind, in which society has secured the means for obtaining a better life through combination of social, economic and institutional processes. The definition of "a better life" may vary from one society to another. Development in all societies must consist of at least the following three objectives:

1. To increase the availability, distribution and accessibility of life-sustaining goods, such as food, shelter, health, security and protection to all members of society;

2. To raise standards of living, including higher incomes, the provision of more jobs, better education and better health, and more attention to cultural and humanistic values so as to enhance not

only material well-being, but also to generate greater individual community and national esteem.

3. To expand the range of economic and social opportunities and services to individuals and communities by freeing them from servitude and dependence on other people and communities and from ignorance and human misery.

Development and Economic growth were used interchangeably for a long time. Although the two are related, they are, however, different. Economic growth can be defined as an increase in the country's productive capacity identifiable by a sustained rise in real national income over a period of years. The differences between growth and development can be outlined as follows.

Development encompasses the total well-being of the individual, a community or a nation, while economic growth is concerned with the increase in per capita earnings of the people making up the nation.

Economic growth is one characteristic of development. It is possible for a country to experience economic growth without becoming developed. A country, for example, may acquire a great wealth from its mineral deposits, but have a low level of health services. This is due to the fact that the wealth goes into the hands of a very small minority who might squander it on luxury goods instead of establishing a viable infrastructure.

Development is concerned with the total person, his/her economic, social, political, physiological, psychic and environmental requirements. If one of these is not fully catered for, development has not been achieved.

The *measurement of development* has presented social scientists with a problem of finding the suitable tools and techniques to do so and of interpreting the results of such measurements. Several suggestions have been presented for measuring development. One line of research has suggested the use of social indicators. The purpose of them is to measure the well-being of the population by examining such factors as health and nutritional status, level of education, housing conditions, and so forth. However, it is easier to calculate GNP, per capita incomes and growth rates. As a result, in most reports these variables are used as indicators of development.

In addition to a rise in per-capita income, economic development implies fundamental changes in the structure of the economy characterized by:

– an increase in the share of industry along with a declining share of agriculture in GNP and an increase in the proportion of people who live in cities rather than the rural areas or villages;

– changes in consumption patterns as people no longer spend all their income on necessities, but move on to consume durables and to leisure-time products and services;

– meeting the needs of the present without compromising the ability of future generations to meet their own needs (sustainability);

– participation by the citizens of the country in the process, as well as the benefit, while economic development and modern economic growth involve much more than a rise in per capita income, there can be no development without economic growth.

The *associations between health and national development* are complex. This interaction is a two-way phenomenon since economic development has an impact both on health and on itself. Improved health has been considered solely a result of economic growth, a part of the product of growth rather than one of its causes. Some development experts have maintained that health should have low priority in development funding and have tried to justify their opinions with comments, such as "only a rich nation can afford the programs to assure its population's health", or "a poor nation cannot afford improved health". The concern of development planners is emphasized by the fact that during the demographic transition, lower death rates are often associated with sustained high birth rates, which results in a rapid population growth. While the supply of labor may increase as a result of improved health and reduced death rates, there may be no corresponding gain in per capita output. Thus, if economic growth is too slow to absorb the additions to the labor force associated with expanded health programs, greater unemployment may result. Thus, improved health in poor societies can be postulated to produce larger populations, greater poverty and ultimately deterioration in health.

However, other development planners and economists are more optimistic regarding the impact of health and nutrition programs

on economic growth. There are three different ways by which improved health programs can accelerate development.

✓ Improved health may increase productivity or efficiency of the labor force leading to greater output and reduced cost per unit of output.

✓ Better health conditions may serve to open new regions of a country for settlement and subsequent development.

✓ Attitudinal changes towards entrepreneurship may be linked to health and nutrition programs. This linkage may stimulate entrepreneurship in poor countries.

It has been apparent that where conditions are worst, relatively simple and low-cost health programs can produce dramatic reductions in disability of the labor force. In these situations, major increments in productivity are readily apparent. For instance, in the Philippines at one time a survey of major enterprises indicated a daily absenteeism rate of 35 percent, attributed largely to malaria. After the initiation of an anti-malaria program the rate of absenteeism was reduced to 2-4 percent, and nearly one-fourth fewer laborers were required for any given task. Although one could argue that economic growth has to accelerate the eradication of poverty, many economists felt that its impact occurred too slowly. In other words, many do not believe in an immediate trickle-down effect of economic growth. Subsequently, a more direct method of poverty reduction, namely the basic needs approach, was advocated. Its aim was the direct fulfillment of basic needs, such as health, clothing, sanitation, shelter, nutrition and education.

UNIT 3

HEALTH AND MEDICAL CARE: AN ECONOMIC PERSPECTIVE

Content

Introduction

Health is defined as "a state of physical mental and social well-being and the absence of disease or other abnormal condition." Economists take a radically different approach. They view health as a durable good, or a type of capital that provides services. The flow of services produced from the stock of health "capital" is consumed continuously over an individual's lifetime. Each person is assumed to be endowed with a given stock of health at the beginning of a period, such as a year. Over the period, the stock of health depreciates with age and may be augmented by investments in medical services. Death occurs when an individual's stock of health falls below a critical minimum level.

Naturally, the initial stock of health, along with the rate of depreciation, varies from individual to individual and depends on many factors, some of which are uncontrollable. For example, a person has no control over the initial stock of health allowed at birth, and a child with a congenital heart problem begins life with a below-average stock of health. However, we learn later that medical services may

compensate for many deficiencies, at least to some degree. The rate at which health depreciates also depends on many factors, such as the individual's age, physical makeup, lifestyle, environmental factors, and the amount of the medical care consumed. For example, the rate at which health depreciates in a person diagnosed with high blood pressure is likely to depend on the amount of the medical care consumed (Is this person under a doctor's care?), environmental factors (Does he/she have a stressful occupation?), and the lifestyle (Does the person smoke or have a weight problem?). All these factors interact to determine the person's stock of health at any point in time, along with the pace at which it depreciates.

Objectives

After studying this unit, you should understand:

✓ the utility analysis in healthcare services;

✓ the conceptual definition of medical care from the economic perspective;

✓ the analysis of the good health production.

Main Content

3.1. Why Good Health? Utility Analysis

Health, like any other durable goods, generates a flow of services. These services yield satisfaction, or what economists call **utility**. Your television set is another example of a durable product that generates a flow of services. It is many hours of programming, or viewing services that your television provides that yield utility, not the set itself. As a product, health is desired for consumption and investment purpose. From a consumption perspective, an individual wants to stay healthy because he/she benefits from an overall improvement in the quality of life. In simple terms, a healthy person feels great and, therefore, is in a better position to enjoy life. The investment element concerns the relation between health and time. If you are in a positive state of health, you allocate less time to sickness and pursue other activities, such as leisure. Economists look at education from the same perspective. Much as a person invests in education to enhance the potential to command a higher

wage, a person invests in health to increase the likelihood of having healthier days to work and generate income.

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

Naturally, each individual chooses to consume that combination of goods and services, including the services produced from the stock of health that provides the most utility. The isolated relation between an individual's stock of health and utility is captured in Fig. 3.1 where the quantity of health, *H*, is measured on the horizontal axis and the level of utility, *U*, is represented on the vertical axis.

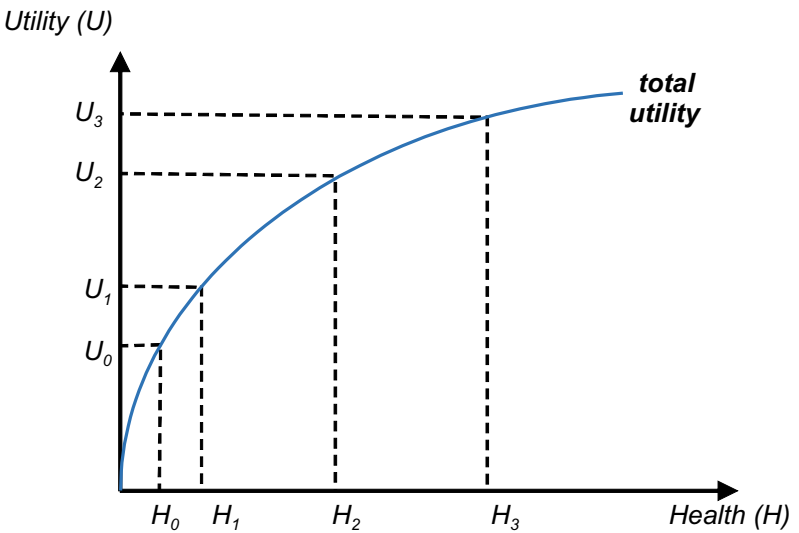


Fig. 3.1. The Total Utility Curve for Health

The total utility curve is upward sloping and depicts the relation between an individual's stock of health and utility. The positive slope indicates that total utility increases as an individual's stock of health improves; the bowed shape of the curve captures the impact of the law of diminishing marginal utility. This law is a fundamental principle of economics stating that each additional improvement in health generates a smaller increase in utility. Note that the increase in health from H_0 to H_1 causes utility to increase from U_0 to U_1, while an equal increase in health for H_2 to H_3 results in a smaller increase in utility from U_2 to U_3.

The positive slope of the curve indicates that an increase in a person's stock of health directly enhances total utility. The shape of the curve is particularly important since it illustrates the fundamental economic principle of the law of diminishing marginal utility. This law states that each successive incremental improvement in health generates smaller and smaller additions to total utility. In other words, utility increases at a decreasing rate with respect to health.

For example, in Fig. 3.1 an increase in health from H_0 to H_1 causes utility to increase from U_0 to U_1, while an equal increase in health from H_0 to H_1 generates a much smaller increase in utility, from U_0 to U_1. In the second case, the increase in utility is less when the stock of health is greater due to the law of diminishing marginal utility. The implication is that a person values a marginal improvement in health more when sick (i.e., when having a lower level of health) than when healthy. This does not mean that every individual derives the same level of utility from a given stock of health. It is possible for two more people to receive a different amount of utility from the same stock of health. The law of diminishing marginal utility requires only that the addition to total utility decreases with successive increases in health for a given individual.

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

$$MU_H = >U/>H \qquad (3.1)$$

where MU_H equals the marginal utility of the last unit of health consumed and represents the change in utility of health (Fig. 3.2).

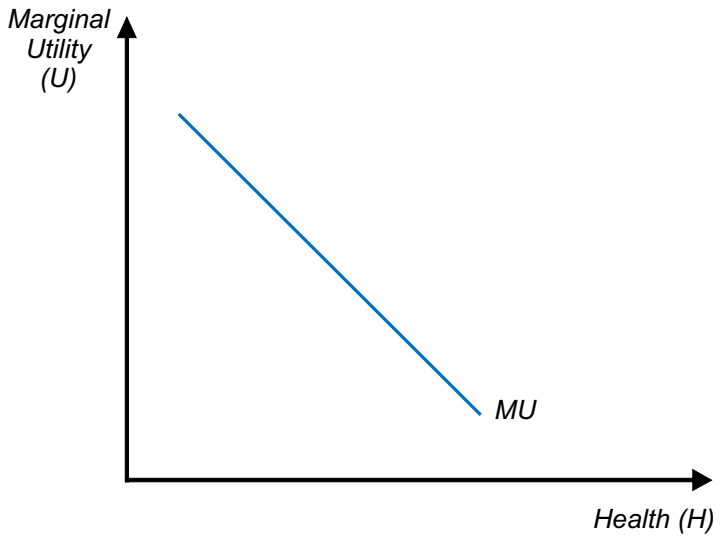


Fig. 3.2. The Marginal Utility Curve for Health

Equation 3.1 represents the slope of a tangent line at each point on the total utility curve. The bowed shape of the total utility curve implies that the slope of the tangent line falls as we move along the curve, or that MU_H falls as health increases. Fig. 3.2 captures the relation between marginal utility and the stock of health. The downward slope of the curve indicates that the law of diminishing marginal utility holds because each new unit of health generates less additional utility than the previous one.

3.2. What is Medical Care and Pharmaceutical Care?

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

Because of their heterogeneous nature, units of medical care are difficult to measure precisely. Units of medical care are also hard to quantify since most represent services rather than tangible products. As a service, medical care exhibits four characteristics that distinguish it from a good: intangibility, inseparability, inventory, and inconsistency. The first characteristic, intangibility means that a medical service is incapable of being assessed by the five senses. Unlike a new car or a new CD, the consumer cannot see, taste, feel, or hear a medical service.

Inseparability means that the production and consumption of a medical service take place simultaneously. For example, when you visit your dentist for a checkup, you are consuming dental services at the exact time the dentist is producing them. In addition, a patient often acts as both a producer and a consumer. Without the patient's active participation, the medical product is likely to be poorly produced. Inventory is directly related to inseparability. Because the production and consumption of a medical service occur simultaneously, healthcare providers are unable to stockpile or maintain an inventory of medical services. For example, a dentist cannot maintain an inventory of dental checkups to meet demand during peak periods. Finally, inconsistency means that the composition and quality of medical services consumed vary widely across medical events. Although everyone visits a physician at some time or another, not every visit to a physician is for the same reason. One person may go for a routine physical, while another may go because he needs heart bypass surgery. The composition of medical care provided or the intensity at which it is consumed can differ greatly among individuals and at different points in time.

The quality of medical care is also difficult to measure. Quality differences are reflected in the structure process, and/or outcome of a medical care. The structural quality is reflected in the physical and human resources of the medical care provider, such as the facilities (level of amenities), medical equipment (type and age), personnel (training and experience), and administration (organization structure). The process quality reflects the specific actions healthcare providers take on behalf of patients in delivering and following through with care. The process quality may include access (waiting time), data collection (background history and testing),

and communication with the patient, diagnosis and treatment (type and appropriateness).

The quality of the outcome refers to the impact of care on the patient's health and welfare as measured by patient satisfaction, loss of working time due to disability, or mortality rate after medical care. Since it is extremely difficult to keep all three aspects of quality constant for every medical event, the quality of medical services, unlike that of physical goods, is likely to be inconsistent.

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

Pharmaceutical care is a complex of organizational, economic, special (medical and pharmaceutical) and social measures aimed at preserving, improving and eliminating physical and, as a result, moral suffering of people with the use of medicinal products and medical devices. Pharmaceutical assistance (PA) is provided regardless of the social and material status of citizens in society, race and nationality, religion, citizenship, age, gender, sexual orientation. As a complex concept, the content and forms of providing pharmaceutical services are constantly transforming in accordance with changes in the external social, economic, scientific and technical environment.

Pharmaceutical aid has a special (pharmaceutical), market, economic and social content. Its constituent elements are the process of providing the population with medicines and medical devices, pharmaceutical care, pharmaceutical ethics and deontology. Pharmaceutical care includes pharmaceutical prevention and diagnostics. Pharmaceutical care is provided at three main levels: health-preserving, life-sustaining, and service.

The life-saving level of pharmaceutical aid is guaranteed by the state and society, provided by the financial institutions of the

national healthcare system, which includes pharmacy as a component of the healthcare system and is determined by the use of medicinal products to preserve human vital functions, e.g. treatment of emergency conditions, pathologies that threaten the life of patients.

In 1990, Charles Hepler and Linda Strand published the first meaningful definition of pharmaceutical care. They wrote: "Pharmaceutical care is the responsible provision of drug therapy to achieve specific outcomes that improve the patient's quality of life." Most European countries rely on this definition in their approach to pharmaceutical care.

Various terms used in Europe, such as "Farmaceutische Zorg" in Belgium, "Farmaceutische Patiëntenzorg" in the Netherlands, "Pharmazeutische Betreuung" in German-speaking regions, "Farmaceutisk/Farmacøytisok Omsorg" in most Scandinavian languages or "Soin Pharmaceutique" in French-speaking regions, mainly refer to pharmaceutical care in terms of the definition of C. Hepler and L. Strand. In addition, there are seemingly heterogeneous concepts with similar meanings, such as "Seguimento Farmacoterapéutico" with the unusual translation of "continuation of pharmacotherapy" in Spain and Portugal and "medication management" in parts of the United Kingdom (UK). In addition to this confusion, there may also be differences in the interpretation of the term "pharmacological care" within the same country or within different institutions (e.g., community or hospital pharmacy).

In accordance with the original American definition, pharmaceutical care is often seen in Europe as a process of optimizing the results of a patient's drug therapy – nothing more and nothing less. The goal of this process is to improve the patient's quality of life (QOL). An optimized medication regimen is not an end in itself, but rather the goal is to improve clinical, economic and/or humanistic outcomes. The question "Who does it?" is relevant to protecting the professionalism of pharmacists, but not necessarily from the patient's perspective.

In most European countries (except the UK), hospital pharmacy services for patients are still underdeveloped. With a few exceptions, most hospital pharmacies and the pharmacists who work in them focus on management issues and preventing medication errors rather than providing pharmaceutical care to identify and manage drug-related problems.

3.3. Production of Good Health

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

Health = H (medical care, technology, profile, lifestyle, socio-economic environment)

where health reflects the level of health at a point in time; medical care equals the quantity of the medical care consumed; technology refers to the state of medical technology at a given point in time; profile captures the individual's mental and physical profile as of a point in time; the lifestyle represents a set of lifestyle variables, such as diet and exercise; socioeconomic status reflects the joint effect of social and economic factors, such as education, income and poverty; and environment stands for a variety of environmental factors, including air and water quality. To focus on the relation between health and medical care, we assume initially that all other factors in the health production function remain constant. Fig. 3.3 depicts this relation, where q is a hypothetical measure of healthcare, holding technology constant, and H represents the level of health. The intercept term represents the individual's level of health when zero medical care is consumed. As drawn, the total product curve implies that an individual's level of health is positively related to the amount of medical care consumed.

The shape of the curve is very similar to that in Fig. 3.3 and reflects the law of diminishing marginal productivity. This law implies that health increases at a decreasing rate with respect to additional amounts of medical care, holding other inputs constant. For example, suppose an individual makes an initial visit and several follow-up visits to a physician's office for a specific illness or treatment over a given period of time. It is very likely that the first few visits have a more beneficial impact on the individual's stock of health than the later visit.

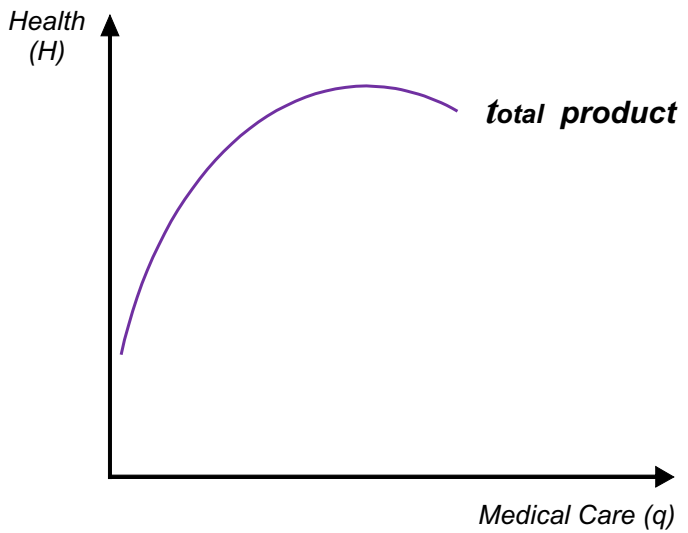


Fig. 3.3. The Total Product Curve for Medical Care

Thus, each successive visit generates a smaller improvement in health than the previous one.

$$MPq = >H/>q \qquad (3.2)$$

where MPq equals the marginal product of the last unit of medical care services consumed. The law of diminishing marginal productivity holds that the marginal product of medical care diminishes as the individual acquires more medical care. A graph of this relationship appears as a negatively sloped curve in Fig. 3.4.

Other variables in the health production function can also be incorporated into the analysis. In general terms, a change in any

one of other variables in the production function alters the position of the MP curve. The MP curve may shift in some instances and/or rotate in others. In the latter case, the curve rotates because the marginal productivity of medical care has changed in response to the change in the other factors.

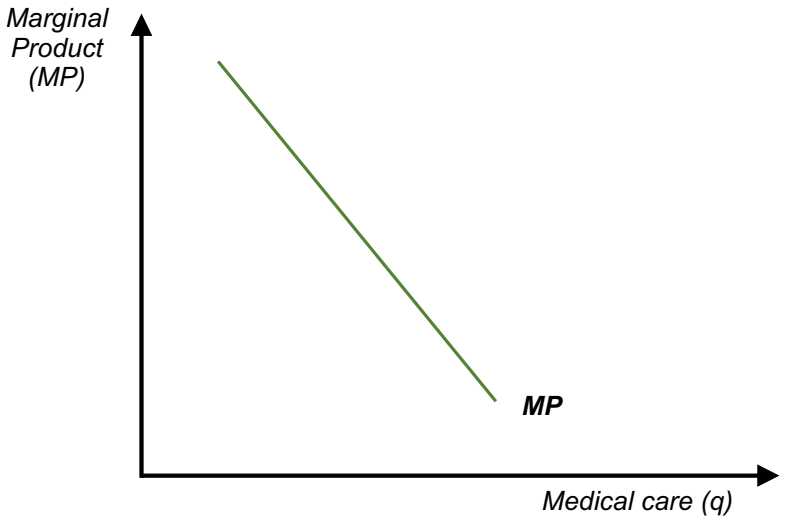


Fig. 3.4. The Marginal Product Curve for Medical Care

New medical technologies have affected all aspects of the production of medical care. In the broadest of terms, examples of new technologies include the development of sophisticated medical devices, the introduction of new drugs, the application of innovative medical and surgical procedures, and most recently, the use of computer-supported information systems. Technological changes can result in the treatment expansion, treatment substitution, or some elements of both. The treatment expansion occurs when more patients are treated with a new medical intuition, which happens when a new technology replaces an older one.

In the context of our health production model, the development and application of a new medical technology causes the total product curve to pivot, and rotate upward because the marginal productivity of each unit of the medical care consumed increases, as illustrated in Fig. 3.5.

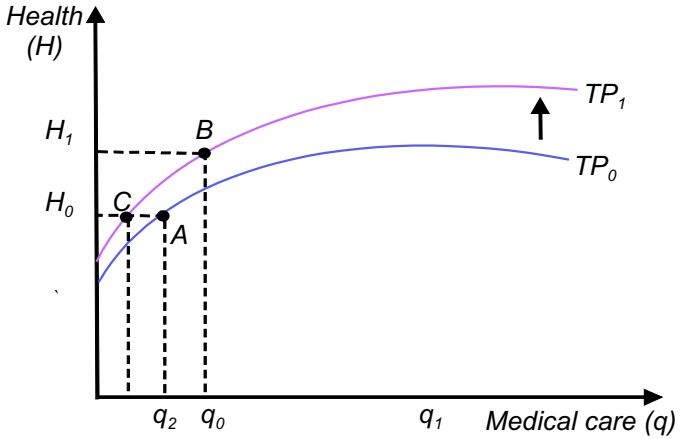


Fig. 3.5. The Effect of Technological Change
on the Total Product Curve for Medical Care

Important! The total product curve rotates upward from TP_0 to TP_1 and each unit of the medical care consumed now generates a greater amount of health.

The movement from point A to point B in Fig. 3.5 illustrates the case, in which the improvement in the medical technology brings about an increase in the amount of the medical care consumed from q_0 to q_1 along with an improvement in health from H_0 to H_1. This movement represents the treatment expansion resulting from the new medical technology. The movement from point A to C illustrates the situation, in which the new technology has no impact on health, but results in less consumption of medical care from q_0 to q_2. In this case, the new technology is cost-saving, all things being equal. It should be noted that the increase in the marginal product of medical care brought about by the medical technology also causes the marginal product curve to shift to the right.

UNIT 4

HEALTH ECONOMIC CARE – PUTTING A PRICE TAG ON HEALTH: COST CONCEPTS, ECONOMIC EVALUATION

Content

Introduction

Economists define cost as the value of resources used to produce a product or services. However, the way these resources are measured can differ. There are two main alternatives with respect to measurement of these resources: financial and economic costing. Financial cost represents actual expenditure on goods and services. Costs are described in terms of how much money has been paid

for the resources used in the project or services. To determine the financial costs of a project, we need to know the price and quantity of all the resources used or the level of expenditure on these goods and services. Economists conceptualize costs in a broader way. They define costs in terms of the alternative uses that have been forgone by using resources in a particular way. These economic or opportunity costs reflect the cost of using resources as these resources are not available for productive use elsewhere. The basic idea is that things have a value that might not be fully captured in their prices. This unit discusses costs in healthcare services and health programs and looks at its implication for economic evaluations.

Objectives

At the end of the unit, the student will be able to:

- ✓ understand the meaning and basis of cost as a concept;
- ✓ be aware of the possibility of using cost concepts to undertake economic evaluation;
- ✓ reconsider the conceptual meaning of opportunity costs;
- ✓ describe direct, indirect, and intangible costs.

Main Content

4.1. Costs Concepts and Types of Costs

It is not difficult in many health programs to identify resources inputs for which little or no money is paid: volunteers working without payment; health messages broadcasts without charge; vaccines or other suppliers donated or provided at large discount by organizations or individuals. Thus, the values of these resources to society, regardless of who pays for them, are measured by opportunity cost. Economic cost then includes the estimated value of goods or services, for which there were no financial transaction, or when the price of a specific good did not reflect the cost of using its productivity elsewhere. The main differences between financial and economic costs are how they relate to:

1. donated goods and services;
2. others inputs with incorrect or distorted prices;
3. valuation of capital assets.

The theory and concept of cost stem from the fact that economic resources are scarce by nature. If it were not for the scarcity of resources, the concept and theory of cost might not exist as such. Scarcity has two sides: the infinite nature of human wants, the finite or limited nature of resources available to produce goods and services.

The cost classification can be done in various ways depending on its nature and specific purpose. There are various types of costs classified into logical groupings. These groups are such that every item of cost can be classified. These classifications of costs make the cost information meaningful. It is of utmost importance to the management of a manufacturing concern (Fig. 4.1).

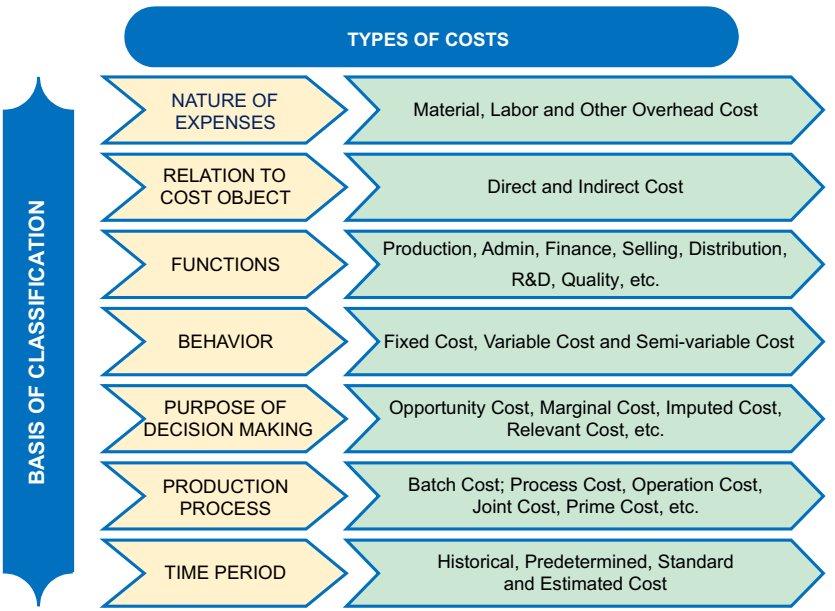


Fig. 4.1. The Cost Classification in Economics

Generally, costs can be considered as ***direct***, ***indirect*** and ***intangible***.

Direct costs are those immediately associated with an intervention, such as staff time, consumables, etc. Indirect costs may include the loss of a patient's job due to treatment. Intangible costs may be things like pain, anxiety, quality, etc. Benefits, however, can

be analyzed in three different ways reflecting the different types of economic analysis used in evaluation. First, benefits can be examined in terms of the immediate (direct) effects on health. These are usually clinically defined units appropriate to the area of study, such as "lives saved", "reduction in tumor size", "change in blood pressure", etc. Second, benefits from an intervention can be considered in more generic terms, such as the impact on general well-being/ happiness/ satisfaction, these are more generally labeled as "utilities". The utility of an intervention to an individual is its benefit. Measures, such as the Quality Adjusted Life Year (QALY) are used to quantify this third way. Benefits might be considered in the same terms as costs, which mean that benefits must be valued in monetary terms by some means.

Whatever kind of economic evaluation may be applied, the costs must be assessed. These are divided into costs borne by the Ministry of Health (for example, drugs and equipment), by patients and their relatives (for example, transportation and food) and by the rest of society (for example, health education).

The costs should be estimated in monetary terms:

1. Direct costs, such as wages, pose little problem.

2. But indirect costs (for example, time spent in the hospital) should have values imputed to them.

3. Costs should also be further subdivided into average, marginal and joint costs, which help to make decisions about how much services should be provided.

4. Capital costs (investment in plant, buildings, and machinery) are also important for due consideration, as discounting and inflation.

If the assessment is carried out from the broadest point of view, from the point of view of society as a whole, then three main categories of costs must be considered.

1. *Health service costs*. They include staff time, medical supplies (including drugs), bed and food services in the case of inpatients, use of capital equipment, and overheads, such as water, heating and lighting. These items may be divided into variable costs, which vary according to the level of activity (for example, staff time) and fixed costs, which are incurred whatever the level of activity (for example, heating and lighting). In the long run, practically all costs become variable since those that are fixed in the short run may be

varied, for example, by opening and closing wards, and by building new hospitals. In economic evaluation, all such health service costs – both fixed and variable – are referred to as direct costs.

2. *Costs Borne by Patients and their Families*. These will include out of pocket expenses, such as travel, and any cost resulting from caring activities undertaken by the family. These are both direct cost items. In addition, there may also be indirect costs (productivity costs), such as income lost because of the absence from work (which is a production loss to society) and any psychological stress experienced by patients, or their families or both.

3. *External costs*. They occur when people not directly involved in a program experience increased costs because of it. In most cases these effects are too small and diffuse to merit inclusion in the analysis, but there may be some occasions when they are large enough to require attention. For example, public health legislation enforcing anti- pollution standards or specifying water purification levels may lead to increase in manufacturing costs and consumer prices (as well as providing health benefits).

How should **costs be valued**? Adequate valuation of costs must consider the following things.

– The costs identified in physical units (such as hours of staff time, hours of the operating theatre use, quantities of drugs and so on) must be valued in monetary terms.

– For most direct cost items market prices will be available.

– Nursing time can therefore be valued at the appropriate hourly rate.

– Medical and surgical supplies can be valued at the prices charged by suppliers.

– Electricity and water can be valued at the appropriate tariffs, and so on.

Strictly speaking, economic evaluation should seek to value all inputs in terms of their opportunity costs, that is, their value in their next best use.

1. These measures what is being given up to use resources in healthcare.

2. Sometimes opportunity costs may diverge from market prices. For example, a nurse would otherwise be unemployed, and then his or her opportunity cost would be zero, and not an hourly wage.

3. For most practical purposes, however, it is usual to use market prices unless there is strong evidence to suggest that they diverge appreciably from opportunity costs.

Indirect costs, for which there are no market prices, pose a more difficult problem of evaluation. Some method has to be used to impute values to them.

– This is known as "shadow pricing", and time costs provide a good example.

– When time is spent in the hospital by a patient, or on caring by a relative, and this displaces work time, it is usual practice to use the relevant wage to value the lost time.

– If it is not work time that is displaced, however, other measures must be used.

Important! Most decisions in healthcare are not concerned with whether or not a service should be provided, or whether or not a particular procedure should be undertaken, but with how much of the service should be provided. That is, should existing levels of provision be expanded or contracted? For example, what family planning services should be made available? This decision requires that attention should be focused on marginal costs, that is, the change in total costs resulting from a marginal change in activity. In the short run, there is often an important difference between the marginal costs of an activity and its average cost, where the average cost is defined as the total cost divided by the total number of units of output. One context, in which the distinction between average and marginal costs is important is in relation to duration of hospital stay of inpatients. Many new procedures have reduced the amount of time necessary for a patient to remain in the hospital and thereby yield cost savings. When evaluating these savings, however, it is important to keep in mind that using average costs per day will generally leads to overestimation of savings since later days of stay are usually cheaper than earlier ones. It is the marginal costs/day that is the relevant measure. Yet another problem of cost measurement arises in connection with joint costs. Often, a single production process can result in multiple outputs. For example, a single chemical analysis of a blood sample can diagnose the presence of many diseases. How should the cost be allocated to each diagnosis? Similarly, within a hospital setting, there are many common services

(e.g., medical records, radiology, operating theatres, laundry, catering, and cleaning) that contribute to a number of specialties. Economic evaluation requires some method for allocating the joint costs of these services to individual programs or procedures. There are several methods, which may be used to do this. Most of them use some physical unit of utilization, such as the number of laboratory tests, hours of the operating theatre use, or square meters of the ward space, to allocate total laboratory and ward cleaning costs.

Capital Costs. Investments in buildings, plant, and equipment that yield a flow of services over a number of years give rise to capital costs. Generally, investment expenditure will be undertaken at the beginning of a project, but the use of items of capital equipment will generate annual capital costs over the lifetime of the asset. These costs have two elements: interest and depreciation.

– Interest costs should be included even if the asset is not acquired with borrowed money because tying up money in an item of capital equipment involves an opportunity cost, that is, interest foregone.

– Depreciation costs arise because of the wear and tear that an asset gets through use and the consequent reduction in the length of its life. (But land is a capital asset that is not assumed to incur depreciation costs).

Sometimes an item of capital expenditure is unique to a particular use and has little or no alternative use value (opportunity cost). In such cases, it is referred to as sunk cost. A hospital building or an item of medical equipment may, for example, have considerable value in its existing use, but little resale value. This can provide a powerful case for continuing to use existing assets instead of undertaking new investments since, in an economic evaluation, sunk costs should not be included among annual capital costs. In practice, this consideration is likely to be more important in the case of major capital developments than of individual procedures.

4.2. Issues in the Measurement of Costs

There are two practical problems in economic evaluation – measuring costs and measuring benefits. Measuring costs appears to be easier than measuring benefits. In almost all organizations there are some attempts to measure costs if only for the purposes

of financial control and accountability. But measuring costs accurately is often difficult, and there are important conceptual and practical problems to overcome. Cost in economics is the opportunity foregone, and in economic evaluation the aim is to measure opportunity cost. This implies that cost estimates produced for the purpose of economic evaluation will not be applicable to all possible other purposes that cost estimates might serve. For example, if a health service manager is interested in the financial implications of the introduction of a new service; or even the maximum effect (in health or welfare terms) achievable for a given expenditure by the health service, a separate analysis of this would need to be commissioned. The justification for the choice of opportunity cost is that it takes into account the costs of all members of society, as they impinge on the social welfare function, and is consistent with the attempt to measure benefits in the same way. Alternative conceptions of cost, which are not consistent with benefit measurement, can lead to illogical conclusions. Nevertheless, if we believe that health services are "under-funded" relative to other parts of the economy we might mean that money in health service hands has a higher value than elsewhere. This could serve as a justification for analyzing the costs only for the health service to achieve health improvement. To measure opportunity cost, we need to know the context, in which choices are made. Good costing exercises start from a clear understanding of how current or potential services operate, what resources are used for particular groups of patients, which are shared, and how the staff spend their time. For example, in one study that was assessing the cost of a long-stay mental health facility, it was found that the staff actually spent most of their time caring for a small group of patients with the most serious problems. This meant that very little direct support was provided for the majority of residents. The actual use of the staff resources, and therefore, the way, in which costs varied between different groups of residents, would not have been observed if costs had been calculated only from accounting data. It is likely that costs would have been assessed as being the same for all residents in the facility. A good understanding of current provision can also help to identify if there is any spare capacity, which would allow the service to be expanded at low cost. It can be important to assemble information about the choices of the

technology and organizational structure available at different scales of provision. It is unusual for it to be possible to assess the costs of new developments accurately without quite detailed knowledge of the technology, management and human skills needed.

Some costs are fixed; some can be changed, but only slowly. Some elements of cost can be easily observed and are obviously related to a particular activity, but others, especially buildings and land, senior staff, equipment and administration, may not vary directly with the level of activity, and it may be difficult to allocate these costs. The simplest approach to calculating costs looks in detail at all the inputs into a service, multiplies by the unit cost of each, and thereby calculates the total cost. This can present an accurate account of the direct costs of a particular service although overhead costs may be hard to allocate. A drawback of this approach is that it does not demonstrate clearly how costs are likely to behave in the event of changes in scale, case mix or technology. There is therefore an argument for trying to estimate cost functions from information on costs and outputs in a larger number of service providers. These data are analyzed using statistical methods to identify how costs vary with the level and mix of output, and to identify the factors that affect costs. It is quite common for costing exercises to use a mixture of approaches since there are usually constraints on access to appropriate data. In some studies, a cost-function approach is used to calculate the unit costs to be applied to activity data.

Costing is not a simple technical exercise – it is too important to leave to accountants alone. Understanding the services provided, as well as the financial data and analysis, are important. Some examples may help. Many health interventions exhibit economies of scale, so that increasing the output may allow a lower cost service to be developed. In these circumstances, the average cost based on current services will overestimate the costs of an expansion. Most emergency services exhibit economies of scale due to the possibility of using the capacity more intensively. An example is neonatal care. Since the need for such services is inherently unpredictable, most centers aim to keep at least one bed free at any time. The proportion of empty beds is therefore lower when there are fewer, larger centers. In cases like these it would be very misleading to cost services at the average if changes in the scale of provision were being contemplated.

Other services are less likely to show significant scale economies, such as palliative care for people dying of cancer, and many parts of primary and secondary care. In such examples, the advantages of centralization are often balanced by the disadvantages to patients of greater geographical distance between home and facility. It may take time to adapt to higher levels of output, so scale economies may not be realized promptly. The technology used may be lumpy (e.g. bits of equipment come only in certain sizes, so that expansion beyond a certain threshold requires a large additional investment). New approaches to provision of certain services may involve a large change in the scale of provision.

Tradition plays a large part in how services are organized. Many patterns of care owe more to historical accident than careful and rational planning. This means that it is important to understand what inputs are really necessary. For example, immunization schedules may be rationalized, grouping a number of vaccines within a single administered dose. Among the reasons for doing this might be a more efficient use of staff time, but it is unlikely that staffing levels of an immunization program will be immediately adjusted. After some time, it might be apparent that there is a little slack in the immunization program than in another program, and the staff might be reallocated. There are several reasons why costs of care for different patients may vary.

1. First, there may be characteristics of the patient that lead to longer hospital stays or more interventions, and therefore, higher costs.

2. Second, habits and traditions in different hospitals can differ, and this can lead to variation in cost. For example, the policy may be to keep all emergency admissions in hospital for at least one day, or alternatively it would be possible to check all patients admitted and discharge those who are not undergoing active examination or treatment. Considering the somewhat limited evidence about the outcomes related to different patterns of care, it may be difficult to implement lower cost (but possibly lower quality) approaches to care.

3. Third, the providers may have different levels of technical efficiency, and so any given service will have a different unit cost. If we are seeking the opportunity cost, in principle, we should be interested

in identifying the lowest feasible cost of providing a given service. Differences that are explained by patient characteristics must be taken into account. It is less obvious how we should treat different clinical policies – normally they vary most where the evidence is weakest, and we often do not know if lower-cost practices reflect greater efficiency or lower quality. In principle, the opportunity costs should not allow any X-inefficiency (technical inefficiency), so we should try to identify the cost in an efficient care provider.

Technically, cost resulting from inefficiency is not a part of opportunity cost. Simply by using the resources efficiently, it is possible to increase welfare. However, if we are convinced that it is impossible to eliminate X-inefficiency within a particular time scale, then it may be appropriate to include some element of inefficiency in the estimates of cost. In this case, in practice these are the minimum costs of providing the program, in the short term at least.

Determining the appropriate concept and cost indicator can be particularly difficult when economic evaluation is carried out as part of clinical trials and studies. Patients recruited into a study are normally heterogeneous, and some variation in costs is likely. Normally, they are not completely typical of patients who are likely to receive the treatment. Large trials normally recruit patients from many centers, and clinical policies and efficiency will also be important. To assess the cost-effectiveness of a new intervention we need to calculate the costs for those patients likely to be provided with the service, in the ways and places they are likely to receive the service. It is important therefore to know how costs vary with such factors as age, sex, disease severity, co-morbidities, case mix and scale of provision. To do this properly, we need sufficiently large samples of patients so that we can understand the differences and calculate confidence intervals to estimate costs. Since little is currently known about the patterns of costs and these factors, it is not yet easy to estimate in advance the sample sizes needed for cost studies, but it is clear that in some cases the variations are large. There has been widespread criticism of the lack of data on confidence intervals in economic evaluation studies, and a range of methods is often applied to estimate them.

Ideally costing studies calculate unit costs of services from a range of settings, but this is not always feasible. When a number of different

providers show very different unit costs, and they are not explained by patient characteristics or treatment effectiveness, an interesting question arises about which assessment should be used. Since opportunity cost is the objective, there is a case for choosing the lowest observed estimate of unit costs, as discussed above. However, differences in unit costs by the institution may not only reflect differences in technical efficiency as this perspective suggests. Since interventions are administered through a given infrastructure, there is a need to match the ideal infrastructure for this particular intervention and the ideal infrastructure for the health system as a whole. This intervention may be efficiently delivered in medium-sized health centers, whereas others are most efficiently delivered in small or large ones. Additionally, health infrastructure as a whole has to balance technical efficiency questions from a health service perspective with patient access costs. Patterns of human settlement do not present standard problems capable of producing a single best solution to health unit size.

On the assumption that the most important determinant of cost variation between units is technical efficiency, some costing studies use data envelopment analysis (DEA) or stochastic frontier analysis, which aim to show costs of the most efficient care providers. Both these approaches aim to estimate cost functions in terms of the lowest observed costs rather than as the average of those observed. DEA is a non-parametric method, and simply joins up the lowest cost observations to describe the function. Since the probability of measurement error is high, using a method that takes this into account has its advantages. Stochastic frontier analysis aims to do this. Using these methods, the relative efficiency of different hospitals can be estimated by comparing observed cost with the lowest observed cost for a comparable provider. A typical indicator of relative efficiency is the ratio of the cost of the service and the cost of the lowest of the services observed.

As with all statistical methods of estimating costs, the concern must be to ensure that differences in case mix are properly controlled for, so that the lowest observed cost is genuinely an example of greater efficiency and not simply the result of easier cases. With this in mind, there are many advantages in frontier methods to estimate cost. The estimate of cost is the lowest for a comparable

provider and should, therefore, contain less X-inefficiency than the average provider. Thus, the frontier estimate can be viewed as being closer to opportunity cost than the average unit cost for all providers. Of course, such methods can only identify relative efficiency since the comparison is with the most efficient observed provider, and not with one that is necessarily efficient in absolute terms.

The costs we are aiming to estimate will usually be associated with adding a new service, or expanding an existing service. Where we are expanding an existing service, whether it is increasing the level of activity within a unit t or extending a service from one set of units to others, the information obtained from cost functions can be very useful. When a cost function is estimated, it can be used to identify costs at higher or lower levels of output, and with different mixes of cases. By comparing the costs at the present level of activity with the costs at the level after implementation of an expanded service, we can obtain estimates of additional (or incremental) cost. The incremental cost is a similar concept to marginal cost, but in this case the change in service volume may not be small.

Where we are adding a new service – which is not yet provided anywhere in the health system – the existing cost data are probably not very useful, and we are likely to be evaluating experimental provision (as where the economic evaluation is attached to a clinical trial – see below), or building up a hypothetical picture of costs. Nevertheless, our interest is still in incremental cost. Whereas, when we are expanding an existing program, economies of scale cause divergence between average and incremental cost, when we are introducing a new program, economies of scope cause this divergence. In principle, a focus on incremental cost is useful since in assessing options we really want to compare differences in costs and benefits between options. When costs are estimated using measures of changes in activity and a vector of (average) unit costs, the estimated costs or savings are likely to be over or under estimates of incremental cost. It is, of course, possible to make adjustments to the unit cost vector to reflect any economies of scale or scope, and therefore, to derive estimates that are closer to incremental cost. If we know the change in output associated with a development, the incremental cost (calculated from a cost function or from a hypothetical model of a new activity) can be used in the cost vector

in place of average cost. There is continuing controversy about the best estimates of incremental costs. In the short run capital costs are not relevant to measuring incremental costs since there will be no change in capital (and other fixed) costs. If the current service has excess capacity, then there may be little or no need to invest in new facilities and equipment, and there may be no need for the additional staff. Under these circumstances, the incremental cost will only include consumables.

However, in the long run all efficient services will adapt capacity to that which is most efficient. Changing the volume of a service will therefore mean that the fixed costs will change in the long run. For this reason, many economists argue that the correct basis for calculating incremental costs includes any changes in capital and other fixed costs. In many cases this means that the short-run AC is a better proxy for long-run incremental cost than the short-run incremental cost.

In circumstances of economies of scale and scope, divergence between average and marginal or incremental cost applies to the long term. For example, where the underlying cause is "lumpy" investment requirements, or "indivisibilities" (units of investment are large), there will be no long-term reconciliation between the two measures of cost. In these circumstances, it is clearer that adjustments for incremental cost have to be made although there is debate over the extent to which long-run economies of scale and scope exist.

When average cost is being used as an estimate of long-run incremental cost, it is important to check that the change in activity is unlikely to lead to a major change in the most efficient technology of provision. For example, if a new universal vaccination program replaces a smaller selective one, the whole organization of the service, and probably the equipment and staff in use, will change. Average costs are unlikely in these circumstances to be a useful basis for estimating the incremental costs of the additional services. In general, for small changes in the volume of a service it is safe to use short-run average cost as a proxy for long-run incremental cost unless the technology is such that there is spare capacity in the current provision, and the service can be efficiently expanded without additional investment.

4.2.1. Sources of Variation in Cost Measures, Confidence Intervals and Assessing Sample Sizes for Costing

There are many reasons why costs vary for the same service in different locations. There is a useful resemblance here with the measurement of the effectiveness of different treatments in clinical trials and studies. The statistical principles for judging the comparative effectiveness of different interventions are widely accepted. Before the start of a clinical trial there is a calculation of the sample that will be needed to give a particular probability of demonstrating a given difference of the effect with a given level of the statistical significance. Clearly, this calculation is dependent on assumptions about the likely distribution of effects, and this assumed variability in effect for any given treatment is one factor in determining the sample size needed.

In the case of clinical trials, it is normal for the basic unit to be the patient. In most studies, patients are allocated to different treatments (using random allocation if feasible), and variations coming from different facilities or staff skills matter little since in each site the patients are allocated at random. A problem arises in cases where randomization has to be by hospital or district rather than by patient since local facilities or skills may play important roles. There can be similar problems in assessing costs. Since costs for a particular patient depend on the disease severity, co-morbidity, hospital size, location and efficiency, it is not clear whether we need a large sample of patients or hospitals to assess the range within which costs are likely to lie.

There is a growing understanding of how costs change as a result of differences in patient characteristics. Many costing studies, particularly in clinical trials, calculate costs for each patient, and this gives data on the degree of variability. Such evidence can give a basis for the calculation of the sample size that will allow costs for each category of patients to be assessed with reasonable reliability. When economic evaluation is being carried out alongside clinical trials or studies, this should be done. Some studies have shown that

the distribution of service use is highly skewed in certain patient groups, especially in mental health and in cases where some patients receive treatment involving high-technology equipment.

It is usually desirable, but not always feasible, to assess unit costs of services from many different hospitals. In terms of interpreting the results of economic evaluation there are two reasons to be interested in understanding variation in cost between facilities. First, it may be that a particular service or intervention is cost-effective only if provided in a low-cost facility. Knowledge of the structure of costs can allow judgements to be made about where such developments should be located. A good example could be hemoglobinopathy screening. Given the large economies of scale in testing, the service is only likely to be cost-effective if testing can be centralized. Second, unless we know the variation in unit costs in different facilities, there is a risk that the assessment of cost-effectiveness reflects the chance that the evaluation was done at a low or at a high-cost location. This is somewhat analogous to drawing conclusions about the efficacy of a new treatment from case reports or small studies.

Since it is often not possible to calculate unit costs for services in more than a few centers, it can be impossible to explore the range of likely costs using conventional statistical methods. It is still useful to present evidence of variation in unit costs, but confidence intervals for cost variation are only possible if it is possible to include in the study data from a large enough range of providers to allow the distribution to be analyzed. But we should remain interested in the consequences of any errors in estimates of cost, and we should try to ensure that strong recommendations reflect our level of confidence in the estimates.

4.2.2. Using Sensitivity Analysis on Costs

While it is desirable where possible to calculate mean costs and confidence intervals around the mean, since variability may be related to the location of services, it is often impossible to do this. In these circumstances it is still desirable to explore the consequences of variation in costs. This is best done by the sensitivity analysis. There are two ways, in which the sensitivity analysis can be used.

First, a range of plausible assumptions can be tested out (such as plus or minus 15 percent), to see if this is likely to affect the conclusions of the analysis. If there is some basis for judging plausible levels of variation, this is appropriate.

An alternative is to start from the other end, and ask the question "What size of variation in cost would be needed to change the conclusions?" If the conclusion remains the same with even large variations in cost, this may be grounds for accepting the results as robust.

4.2.3. Costing in Economic Evaluation

The normal approach to calculating costs in economic evaluation is to estimate the number of cost-generating events for each patient, and to multiply this matrix of different events for different patients by a vector of unit costs. As suggested above, this unit cost vector may be calculated using a range of methods, from accounting or budget data or estimates of cost functions. In many cases simple approaches have been considered to be adequate, and most studies do not take into account changes in costs with time or technical progress. It was argued at the start of this chapter that costing requires understanding of circumstance, as well as technique. It may be quite acceptable to assume that costs for a particular service will remain stable over time. Equally, there are some instances, in which such an assumption leads to serious errors. For example, in most surgeries in industrialized countries, the length of hospital stays has fallen consistently, and this trend seems likely to continue and to be capable of exploitation in other countries. Failing to take this into account may lead to overestimates of the costs of surgical options in the future. New technologies may reduce in price over time, and may be the subject of learning, suggesting that health workers' skills may develop in such a way that they use the technology more efficiently. Patients and potential patients learn more about the service and how to use it, contributing to reduced costs. For example, a new technology, such as the treatment of bed nets with insecticide to combat malaria, may require aggressive marketing at first, but rely on word of mouth later once its uptake has reached high

levels. Drugs become much cheaper when patents expire. Costing studies should take all these factors into account.

Costing cannot be an exact science, but costs estimated using sensible approaches by people who are well informed about the context are more likely to reflect the true foregone opportunities.

4.3. Economic Evaluation

Economic evaluation may be based on the viewpoint of an individual patient, the hospital, the government, or the society at large. Hence, it is important to determine at the beginning from whose viewpoint an economic evaluation is to be carried out. The broadest viewpoint is that of society, as this will include all the costs and benefits. Adopting this approach has two main implications that distinguish it from approaches with more limited perspectives. Firstly, it usually involves measuring and evaluating items that do not have market prices attached to them, such as the time costs that patients incur when undergoing treatment and recuperating.

Secondly, it means that certain costs, or cost savings, or both, should not be included in the evaluation since they are transfers from one sector to another rather than a net cost to society, e.g. free healthcare.

Characteristics of Economic Evaluation/ Analysis

First, it deals with both the inputs and outputs, sometimes called costs and consequences, of activities. It is the linkage of costs and consequences, which allows us to reach our decision. Second, the economic analysis concerns itself with choices. Resource scarcity, and our inability to produce all desired outputs, necessitates that choices must be made in all areas of the human activity. These choices are made on the basis of many criteria, sometimes explicit, but often implicit. The economic analysis seeks to identify and make explicit set of criteria, which may be useful in deciding among different uses of scarce resources. Economic evaluations:

1. always compare any healthcare program with an alternative, for example, no treatment or routine care;

2. always measure the benefits produced by all alternatives compared;

3. always measure the cost of any program.

The above characteristics of economic evaluation/analysis lead us to define economic evaluation as a comparative analysis of alternative options in terms of both their costs and consequences.

Therefore, the basic tasks of any economic evaluation are

1. to identify,
2. to measure,
3. to assess,
4. to compare the costs and consequences of the alternatives under consideration.

Economic evaluation of healthcare programs aims to aid decision-making with their difficult choices in allocating healthcare resources, setting priorities and forming health policy. But it might be argued that this is only an intermediate objective. The real purpose of doing economic evaluation is to improve efficiency: the way inputs can be converted into outputs (saving life, health gain, improving the quality of life, etc.).

The choice of which medical care to provide depends on what economists call the allocative efficiency. This means that we strive to maximize benefits (regardless of how we choose to measure it) depending on the resources available. So, from a fixed resource we aim to get as much out of a range of healthcare programs as possible. This will mean that we will need to compare very different interventions, for example, health promotion recommendations aimed at quitting smoking, and the prescription of the drug Relenza, and the procedure for removing an ingrown toenail. Thus, allocative efficiency is about finding the optimal mix of services that deliver the maximum possible benefit in total. Resources will be directed to interventions that are relatively efficient at converting inputs into health benefits and away from those that require larger input for relatively low health gain. This approach may be constrained by certain equity considerations, to ensure that certain groups do receive healthcare.

The choice of how to provide healthcare is about what economists call technical efficiency. This means that we might strive for minimum input for a given output. For example, if we have decided that performing tonsillectomies on children is worthwhile, part of an allocative efficiency, then we may need to examine the efficiency of how we do this. So, if the output we wish to achieve is

to remove the child's tonsils, then we might choose between, say, a day case procedure or an inpatient stay. This is an issue of technical efficiency since the output or "outcome" is fixed, but the inputs will differ depending on which policy we adopt. The day case approach may perhaps require more intensive staff input and more follow-up outpatient visits. If this was the case, then inpatient tonsillectomy may be the more technically efficient strategy.

Thus, with any given healthcare program an economic evaluation aims to clearly identify the total amount of resources consumed by a particular program and the total benefits generated by that program. Drummond defines economic evaluation as "the comparative analysis of alternative courses of action in terms of both their costs and consequences." It differs from other forms of analysis since it considers both costs and consequences and is comparative. Evaluation needs to be comparative as an intervention can only be labeled as good or bad relative to some benchmark or alternative even if this alternative is a "do nothing" strategy. If an evaluation is not comparative and does not consider both costs and consequences, then it is only a partial evaluation. It is a description of either the costs or the benefits of one intervention in isolation. This is most uninformative since it is one-dimensional and without a context by which to judge relative performance. If both costs and consequences are considered, but no comparator is provided, then the study is again only a partial evaluation described as a cost-outcome study. It lacks context and is of limited use. If alternatives are compared, but only in terms of costs or benefits and not both, then again the study only provides a partial evaluation and can be labeled an effectiveness study or a cost analysis. It would be comparative, but only across one dimension. Hence, an economic approach can be considered a full evaluation technique.

Whatever the approach, the same three-stage process for the assessment of all costs and benefits can be applied. All relevant cost and benefit variables must be:

- identified,
- quantified, and
- valued.

At the start of an evaluation, it must be determined which costs and benefits are sufficiently important to merit inclusion in the study.

This should be separate from the measurement stage so as to avoid the study being entirely data driven (i.e. the more intangible consequences of an intervention might be considered equally important). The identification of relevant benefits and costs will define the variables in the study. These can be broadly classified into changes in resource use, changes in productive output and changes in health state.

The next stage is to measure changes in these variables brought about by the intervention in question. Often, it is important that this is done before evaluation as it is necessary to know the magnitude of gains or losses before values can be attached. Presenting variables in terms of "natural" quantities or frequencies (i.e. hours worked or clinical units) can also be very useful in terms of generalizability. Others can use these data and apply values relevant to their own setting (i.e. different cost structures or health values).

The differential timing of costs and benefits must also be considered in an evaluation. The effects of health treatments do not always occur at the same point in time. Costs may be incurred today, but the benefit may not arrive until next year (i.e. preventive treatments, health promotion), part of this future benefit might be that future costs will be avoided. N$100 spent today may not have the same value as N$100 spent next year because of inflation; interest on savings and a positive rate of time preference. People may just prefer to have N$100 in their pocket today rather than N$100 in a week or a month or a year because it offers them more choices. This can be incorporated into economic evaluation by the notion of discounting future costs and benefits to their present-day value.

4.4. Types of Economic Evaluation

The different ways of looking at benefits combined with the cost analysis represent different techniques of economic evaluation. The basic types of economic evaluation are cost-effectiveness analysis (CEA), cost minimization analysis (CMA), cost utility analysis (CUA) and cost benefit analysis (CBA). When to apply to each of the above methods will depend on the nature of the issue being addressed, which may be a choice between alternative clinical strategies for the condition: timing of an intervention; settings for

care; types and skill-mix of personnel proving care; programs for different conditions; or other ways to improve health.

The basic elements of economic evaluation are given in Fig. 4.2.

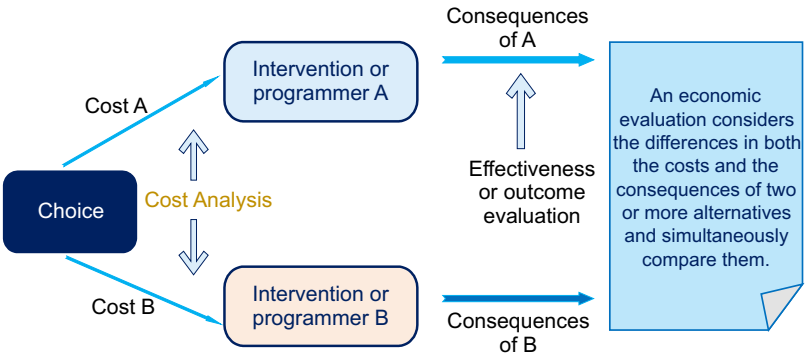


Fig. 4.2. The Basic Elements of an Economic Evaluation

The four methods used in economic evaluations are presented in Fig. 4.3.

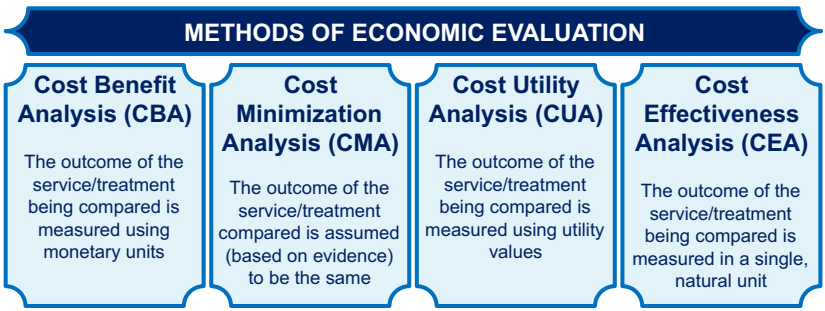


Fig. 4.3. Methods of Economic Evaluation

4.4.1. Cost-Effectiveness Analysis

When different healthcare interventions are not expected to produce the same outcomes both the costs and consequences of the options need to be assessed. This can be done by cost-effectiveness analysis, whereby the costs are compared with outcomes measured in natural units, for example, per life saved, per life year gained, and pain or symptom free day. Many cost-effective analyses rely

on existing published studies for effectiveness data as it is often too costly or time consuming to collect data on costs and effectiveness during the clinical trial. Where there is uncertainty about the costs and effectiveness of procedures, the sensitivity analysis can be used; it examines the sensitivity of the results to alternative assumptions about key variables. CEA is concerned with technical efficiency issues, such as: What is the best way of achieving a given goal? or What is the best way of spending a given budget? Comparisons can be made between different health programs in terms of their cost-effectiveness ratios: cost per unit of effect. According to CEA, effects are measured in terms of the most appropriate one-dimensional natural unit. So, if the question to be addressed was "What is the best way to treat renal failure?" Then the most appropriate ratio for comparing programs could be "cost per life saved". In addition, if we wanted to compare the cost-effectiveness of Down syndrome screening programs, the most appropriate ratio might be "cost per Down syndrome fetus detected". The advantages of the CEA approach are:

1. It is relatively straightforward to carry out.

2. It is often sufficient for addressing many questions in healthcare. However, it is not comprehensive. The outcome is one-dimensional in this analysis, but often health programs generate multiple outcomes.

3. For example, in the Down syndrome screening, the fetus detected is one outcome, but miscarriages avoided might be another very relevant outcome measure, especially if, say, blood testing is being compared to amniocentesis. But this cannot be incorporated into this form of analysis. Thus, CEA not only assumes that the outcome of the health program is worthwhile *per se*, but also that it is the most appropriate measure. A further problem with CEA is comparability between very different health programs. Cost per fetus detected may be a useful way to compare the efficiency of blood testing versus amniocentesis, but how would these be compared to, say, drugs aimed at reducing cholesterol. Health programs with different aims cannot be compared with one another using CEA: cost per unit reduction in cholesterol cannot meaningfully be compared with fetus detected. Hence, CEA is useful when comparing programs within like areas, where common "currencies" can be used.

If the outcomes of alternative procedures or programs under review are the same, then attention can focus upon the costs in order to identify the least cost option. Then, the method of evaluation will be cost-minimization analysis. If, however, the outcomes are not expected to be the same, then both the costs and consequences of alternative options need to be considered. The cost-effectiveness analysis is one method of economic evaluation that allows this to be done.

Measures of Effectiveness. To perform the cost effectiveness analysis, it is necessary to have suitable measures of effectiveness. These will depend on the objectives of the particular interventions under review. In the cost effectiveness analysis, however, measures of effectiveness should be defined in appropriate natural units and, ideally, expressed in a single dimension.

Common measures used in several studies have been "lives saved" and "life years gained". Thus, Boyle and colleagues, in their study of neonatal intensive care of very low birth weight babies, measured effectiveness in terms of mortality rates at the time of discharge of newborn infants from hospital. Their study compared two periods – one before the introduction of neonatal intensive care, and another after its introduction – and measured cost effectiveness in terms of additional costs per life saved. Several other measures of effectiveness have been used by different researchers. They have included the number of pain or symptom free days resulting from alternative drug regimens in the treatment of duodenal ulcers; and the number of episodes of fever cured and deaths prevented in the treatment of chloroquine resistant malaria in African children. Most of the abovementioned studies express effectiveness in terms of a single dimension and thereby permit a direct comparison between alternative procedures in terms of their marginal cost per unit of outcome. Sometimes, however, the alternatives under examination have multiple outcomes. Nonetheless, many of these choices can be dealt with in the cost-effectiveness analysis framework. Thus, if one procedure emerges as less costly and of equal or greater effectiveness than all the other options on each dimension of effectiveness, it is clearly the most cost-effective option. For example, the comparison of day surgery with overnight inpatient care for cataract surgery, measured outcomes in terms of the number of both operative

and postoperative complications, and in terms of visual acuity of patients three to six days and 10 weeks to six months after surgery. The patient satisfaction was also elicited through a questionnaire. As day surgery emerged as the more effective option on practically all of these effectiveness measures, and was subsequently less costly, the evidence suggests that it is the preferred option. One argument for carrying out analysis in this way, that is not always seeking to combine outcome measures into a single unit, is that the variations across a number of dimensions are made clear to decision makers rather than being concealed within an aggregate measure. This can sometimes permit more informed decision-making.

4.4.2. Costs-Minimization Analysis

The cost-minimization analysis (CMA) is an appropriate evaluation method to use when the case for an intervention has been established, and the programs and procedures under consideration are expected to have the same or similar outcomes. In these circumstances, attention may focus on the cost side of the equation to identify the least costly option.

Cost-minimization:

1. it is concerned only with technical efficiency;
2. it can be regarded as a narrow form of the cost-effectiveness analysis;
3. evidence is given on the equivalence of the outcomes of different interventions; and as outcomes are considered to be equivalent, no different decisions can be made on the basis of costs.

The advantages of cost minimization analysis are:

– It is simple to perform, it requires costs to be measured, but only that outcomes can be shown to be equivalent.
– It avoids needlessly quantifying data.

The disadvantages are:

– It can only be used in narrow range of situations.
– It requires that outcomes be equivalent.

An example of CMA can be comparison of two programs involving minor surgery for adults. Both programs achieve the desired result, and from the analysis of the effectiveness data they do not differ from each other in any significant way, except that one

requires hospitalization for at least one night, while the other (a day surgery program) does not. If we identified the common outcome of interest – operations successfully completed – we would find that it could be achieved to the same degree (i.e. identical number of surgeries) in either program, though at different costs. The economic evaluation is then essentially a search for the least cost alternative. An analysis like this is often called the cost-minimization analysis. We might also be interested in the distribution of costs (e.g. in this case to what extent does the day-surgery program shift costs to the patient), but our principal efficiency comparison will be made on the basis of cost per surgical procedure.

Discounting Benefits (in Cost-Effectiveness Analysis)

Costs incurred at different points in time need to be "weighted" or discounted to reflect the fact that those that occur in the immediate future are of more importance than those that accrue in the distant future. This raises the question "Should the benefits or effects of alternative procedures also be discounted?" When answering this question, economists disagree. If zero discounting had been adopted (discounting was not applied), the main consequence would have been a change in the relative economic efficiency of the various procedures. Using a positive discount rate means that projects with long lasting effects receive lower priority. If a positive rate is replaced by a zero rate, procedures, such as neonatal care, which lead to benefits over the recipient's entire future lifetime, will become relatively more cost effective. From a practical point of view, it is probably fair to say that, although the arguments in favor of using a zero discount rate for benefits have a strong intellectual basis and may receive empirical support in the future, it would be too hasty to recommend abandoning positive rates in economic evaluations. In general:

1. The cost-effectiveness analysis is a form of economic evaluation, in which the costs of alternative procedures or programs are compared with the outcomes measured in natural units, for example, cost per life year saved, cost per symptom free day. Effectiveness data are collected based on economic evaluations conducted in parallel with clinical trials. In the absence of specialized trials, researchers need to build on the existing published work.

2. The sensitivity analysis should be applied when there is uncertainty about the costs and effectiveness of different procedures. This investigates the extent to which the results are sensitive to alternative assumptions about key variables.

3. There is a debate among economists about whether benefit indicators should be "time discounted" in the same way as costs. If they are not, projects with long lasting effects will become relatively more cost effective, for example, maternity services and health promotion. But it will be probably wrong to recommend this as a standard practice.

4.4.3. Cost-Utility Analysis

The cost-utility analysis (CUA) is concerned with technical efficiency and allocative efficiency (within the healthcare sector). It can be thought of as a sophisticated form of CEA since it also makes comparisons between health programs in terms of cost-effect ratios. However, CUA differs in the way it considers effects. CUA tends to be used when quality of life is an important factor involved in the health programs being evaluated. This is because CUA combines life years (quantity of life) gained as a result of a health program with some judgment on the quality of those life years. It is this judgment element that is labeled utility. Utility is simply a measure of preference where values can be assigned to different states of health that represent individual preferences. This is done by assigning values between 1.0 and 0.0 where 1.0 is the best imaginable state of health (completely healthy) and 0.0 is the worst imaginable (perhaps death). States of health may be described using many different instruments, which provide a profile of scores in different health domains. For example, EuroQol EQ-5D simplifies health into just five domains (such as mobility, self-care, usual activities, pain/discomfort and anxiety/depression).

✓ Each domain is given a score from 1 to 3.

✓ Thus, the health profile would read 1 for the best scores in all domains and 3 for the worst.

This approach of using utility is not restricted to similar clinical areas, but can be used to compare very different health programs in the same terms. As a result, "cost per QALY gained" league tables

are often produced to compare the relative efficiency with which different interventions can turn resources invested into QALYs gained. It is possible to compare surgical, medical and health promotion interventions with each other. Comparability then is the key advantage of this type of economic evaluation. For a decision-maker faced with allocating scare resources between competing claims, CUA can be very informative. The key problem with CUA is the difficulty of deriving health benefits.

When should CUA be used?

Below there are a number of situations, in which CUA can be used:

1. when health-related quality of life is the important outcome. For example, in comparing alternative programs for the treatment of arthritis, no program is expected to have any impact on mortality, and the interest is focused on how well different programs will improve the patient's physical function, social function, and psychological well-being;

2. when the program affects both morbidity and mortality and we wish to have a common unit of outcome that combines both effects. For example, treatments for many cancers improve longevity and long-term quality of life, but decrease the quality of life during the treatment process;

3. when the programs compared have a range of different kinds of outcomes and we wish to have a common unit of output for comparison. For example, if a health planner who must compare several disparate programs applying for funding, such as expansion of neonatal intensive care, a program to locate and treat hypertension, and a program to expand the rehabilitative services provided to post-myocardial infarction patients;

4. when we wish to compare a program to others that have already been evaluated using cost-utility analysis.

When should CUA not be used?

Below there are a number of situations, in which CUA may not be used:

1. when only intermediate outcome data can be obtained. For example, in a study on screening employees for hypertension

and treat them for one year, the intermediate outcomes of this type cannot be readily converted into QALYs for use in CUA;

2. when the effectiveness data show that the alternatives are equally effective in all respects of importance to consumers (e.g. including side-effects). In this case, the cost- minimization analysis is sufficient; CUA is not needed;

3. when the effectiveness data show that a new program is dominant; that is, the new program is both more effective and less costly (win-win). In this case, no further analysis is needed;

4. when the extra cost of obtaining and using utility values is judged to be in itself not cost effective. This is the case above in points (ii) and (iii). It would also be the case even when the new program is costlier than the old, if effectiveness data show such an enormous superiority for the new program, the incorporation of utility values could almost certainly not change the result. It might even be the case with a program that is costlier and only somewhat more effective, if it can be argued that the incorporation of any utility values will show the program to be overwhelmingly cost-effective.

Measuring Quality

Measuring a person's quality of life is difficult. Nonetheless, it is important to have some means to have for doing so since many healthcare programs are concerned primarily with improving the quality of a patient's life rather than extending its length. For this reason, various quality of life scales has been developed in recent years. The Nottingham health profile is one quality of life scale that has been used quite widely in Britain. This comprises two parts:

1. The first measures health status by asking for "yes" or "no" responses from patients to a set of 36 statements related to six dimensions of social functioning:

- ✓ energy,
- ✓ pain,
- ✓ emotional reactions,
- ✓ sleep,
- ✓ social isolation,
- ✓ physical mobility.

These responses are then "weighted", and a score of between 0 and 100 is assigned to each dimension.

2. The second part asks about seven areas of performance that can be expected to be affected by health:

- ✓ employment,
- ✓ looking after the house,
- ✓ social life, home life,
- ✓ sex life,
- ✓ hobbies,
- ✓ holidays.

The Nottingham health profile has been applied, for example, in studies of heart transplantation, rheumatoid arthritis and migraine, and renal lithotripsy.

Other widely used measures include the sickness impact profile and the quality of well-being scale. Recently, a new outcome measure, the Sf-36 health survey questionnaire, has been gaining popularity. After testing it on 1980 patients in two general practices it is considered to be a promising measure, which is "easy to use, acceptable to patients, and fulfils stringent criteria of reliability and validity". Although all of these scales embody some form of a scoring scheme, they do not usually generate a single quality of life score. This means that, although they are of considerable value in assessing the outcomes of interventions in the case of particular diseases or disabilities, they cannot be used to compare outcomes between different programs. To do this, generalizable measure of quality is necessary. One of the earliest measures to be developed – and one which has subsequently been used widely to calculate QALYs – is the Rosser index.

Rosser Index described the health status in terms of two dimensions: disability and distress. The states of illness are classified into eight categories of disability and four categories of distress. By combining these categories of disability and distress 32 (8 times 4), different states of health were obtained. Rosser then interviewed 70 respondents (a mixture of doctors, nurses, patients and healthy volunteers) and, by using psychometric methods sought to establish their views about the severity of each state relative to other states. The final results of this exercise were expressed in terms of a numeric scale extending from 0 = dead to 1 = perfect health. With this classification system it becomes possible to assign a quality-of-life score to any state of health as long as it is placed in an appropriate disability or distress category.

Quality – Adjusted Life – Years (QALY)

One of the features of conventional CUA is its use of the QALY concept. Results are reported in terms of cost per QALY gained. QALYs:

– Combine life years gained with a measure of the quality of those years.

– Quality is measured on a scale of 0 to 1, with 0 equated to being dead and 1 equated to the best imaginable state of health.

– Combine all dimensions of health and survival into a single index.

$$CU\ ratio = \frac{Cost\ A - Cost\ B}{QALY\ A - QALY\ B}.$$

QALY Concept. The advantage of QALY as a measure of the health outcome is that it can simultaneously capture gains from reduced morbidity (quality gains) and reduced mortality (quantity gains), and combine these into a single measure. Moreover, the combination is based on the relative desirability of different outcomes.

The QALY approach, which forms a key part of most cost-utility analyses, has been the subject of some criticism. It has been accused of discriminating against elderly people, making illegitimate interpersonal comparisons, disregarding equity considerations, and introducing bias into quality-of-life scores. Rival measures that are claimed to be sound theoretically, such as "healthy years equivalents" (HYEs), have also been put forward. It has been claimed that under most assumptions QALYs and HYEs will lead to identical project rankings. Amid all this debate it is as well to bear in mind that decisions have to be made about the allocation of resources and the cost-utility analysis is probably the most sophisticated form of the economic evaluation currently available. However, the technique and interpretation of research findings should recognize that the cost utility-analysis is still at a fairly early development stage and treat it accordingly.

Important! The Disability-Adjusted Life Year (DALY Concept) is a measure akin to QALY in aggregating survival and the quality of life effects, but normally advanced as a method of estimating the burden of illness associated with a disease, rather than the cost-effectiveness of healthcare interventions.

4.4.4. Cost-Benefit Analysis

The cost-benefit analysis is the most comprehensive and theoretically sound form of economic evaluation and it has been used as an aid to decision-making in many different areas of the economic and social policy in the public sector for more than fifty years.

Important! The cost-benefit analysis (CBA) estimates and totals up the equivalent money value of the benefits and costs to the community of projects to establish whether they are worthwhile. These projects may be dams and highways or can be training programs and healthcare systems. The main difference between the cost-benefit analysis and other methods of economic evaluation that were discussed earlier in this series is that it seeks to place monetary values on both the inputs (costs) and outcomes (benefits) of healthcare. Among other things, this enables the monetary returns on investments in health to be compared with the returns obtainable from investments in other areas of the economy. Within the healthcare sector itself; the attachment of monetary values to outcomes makes it possible to say whether a particular procedure or program offers an overall net gain to society in the sense that its total benefits exceed its total costs. The cost-effectiveness and cost-utility analysis do not do this since they measure costs and benefits in different units. CBA requires program consequences to be valued in monetary units, thus, enabling the analyst to make a direct comparison of the program's incremental cost with its incremental consequences in commensurate units of measurement, whether in Birr, dollars, or pounds. CBA compares the discounted future streams of incremental program benefits with incremental program costs; the difference between these two streams is the net social benefit of the program. In simple terms, the goal of the analysis is to identify whether the program's benefits exceed its costs, a positive net social benefit indicating that the program is worthwhile. CBA is a full economic evaluation because the program outputs must be measured and valued. In many respects CBA is broader in scope than CEA/CUA. Since CBA converts all costs and benefits to money, it is not restricted to comparing programs within healthcare, but

can be used to inform resource allocation decisions in the sectors of the economy. CBA is broader in scope and able to inform questions of allocative efficiency as it assigns relative values to health and non-health-related goals to determine which goals are worth achieving, given the alternative uses of resources, and thereby determining which programs are worthwhile.

1. Both costs and benefits are assigned a monetary value. The benefits of any intervention can then be compared directly with any costs incurred. If the value of benefits exceeds the costs, then it is potentially worthwhile to carry that intervention out.

2. It is concerned with allocative efficiency.

3. It is concerned with the question "Is a particular goal worthwhile?" It can answer questions, such as "Should extra money be used for heart transplants or improving housing?".

4. The method requires that all resources and benefits generated by an intervention need to be assigned a monetary value. Therefore, it needs to cost things that have no market value, i.e., changes in health, quality of life, length of life, pain, etc.

Assessment methods are:

- willingness to pay (WTP);
- human capital approach (HCA).

The net welfare gain or net value of a project *X (NVX)* is equal to:

$$NVX = WTPX - WTY$$

where *y* refers to the next best alternative project if the latter cannot be defined.

$$NVX = WTPX - WTPXi$$

where *WTPXi* refers to society while WTP is for the inputs used alternatively in the economy. If *NVX* is positive then, project *X* may be undertaken. When several projects compete with each other, the one with the highest *NV* needs to be selected in order to maximize welfare. This shows the CBA for projects that have benefits or costs in the current period. It is evident that projects may also entail future benefits and future costs. Some modifications in the calculation of the net value will be required in this case. Note that individuals prefer a net value of N$*1* received now to N$*1* in the future. It follows that one cannot simply add up benefits or costs that

are related to different points in time. A social discount rate denoted as *r* will enable us to add up a stream of net benefits, namely, N$*1* in year one will be worth N$*1tr* in year two, N$$1tr^2$ in year three, etc.; conversely, N$1 in year two is worth N$*(1/1tr)* in year one, N$*1* in year three is worth N$$(1/1tr)^2$ in year one, etc. The value in year one of a naira received or paid in the future is called the present value of that naira. Making use of the social discount rate *r*, we can calculate the net present value (NPV) of a project:

$$NPV = [(Bt - Ct) / (1+r) - 1]$$

where *B* and *C* refer to benefits and costs, and *t* is the time index.

Bt is equal to the WTP for the *nth* project at time *t*, while *Ct* has to be understood as the benefits forgone in the period *t*. Note that if $NPV > 0$, the society's welfare will increase; hence, the project can be adopted. If several projects are competing with each other, the one with the highest *NPV* should be chosen.

UNIT 5

SOCIAL MARKET AND HEALTHCARE MARKET

Content

Introduction

The stock of health can be treated as a durable good that generates utility and is subject to the law of diminishing marginal utility. This means that each incremental improvement in health generates successively smaller additions to total utility. Medical services are an input in the production of health since a person consumes medical care services for the purpose of maintaining, restoring, or improving health. However, the law of diminishing marginal productivity causes the marginal improvement to health brought by each additional unit of medical care consumed to decrease.

From this discussion, it follows that medical care indirectly provides utility. Specifically, medical care helps to produce health, which, in turn, generates utility. Consequently, utility can be specified as a function of the quantity of medical care. The shape of the total utility curve indicates that utility increases at a decreasing rate with respect to medical care, or that medical care services are subject to diminishing marginal utility. Marginal utility decreases because each successive unit of medical care generates a smaller improvement in health than the previous unit (due to the law of diminishing marginal productivity) and each increase in health, in turn, generates a smaller increase in utility (due to the law of diminishing marginal utility). This unit deals with the analysis of demand for medical care using the traditional demand analysis.

An implicit assumption for the study of the determinants of demand for medical services and the uncertainty surrounding health needs was that individuals had well-defined and well-informed preferences for health and healthcare, and that they made their consumption and risk-reduction decisions rationally. But this assumption, which is basic to the most economic analysis, must be relaxed to some degree in the analysis of the supply of medical services since it is clear that, as well as providing operational services (that is, injections, surgery, and the like), one of the primary roles of a medical care worker is the provision of information that affects the demand for services. The implications of this connection for the level and quality of care can be significant and lead us to question the efficiency of a market-determined allocation of resources in the healthcare sector. This topic considered the motivation and behavior of physicians and other healthcare sector workers, the institutions within which they operate, and the resource allocation outcomes that alternative financial structures and other incentives might yield. Of particular concern is the efficiency of health service production, the appropriateness of the services, and the allocation of physicians and other labor to rural and urban areas. The analysis is essentially one of the equilibria between supply and demand. However, the special position of medical care providers in relation to their patients means that demand may be in some sense a function of the behavior of suppliers. The interest is particularly in the extent to which physicians can induce consumers

to purchase more medical care than they would if they were fully informed about its effects, and the impact such strategies might have on market responses to supply shocks and the effectiveness of government price interventions.

Objectives

After studying this unit, you should be able to understand:

- ✓ the utility-maximizing rule
- ✓ the demand for medical care services.
- ✓ the elasticity of demand for medical care services
- ✓ the analysis of inputs into the production of health
- ✓ the short-run production analysis of healthcare services.
- ✓ the elasticity of the long-run production analysis of healthcare services

Main Content

5.1. The Utility-Maximizing Rule

Taking into account market prices at a point in time, consumers must decide which combination of goods and services, including medical care, to purchase with their fixed income. According to the microeconomic theory, each consumer chooses the bundle of goods and services that maximizes utility. The consumer utility is maximized when the marginal utility gained from the last naira spent on each product is equal across goods and services purchased. This condition is known as the utility-maximizing rule, and it basically states that the total utility reaches its peak when the consumer receives the maximum "bang for the buck" in terms of the marginal utility per naira of income from each and every good. In mathematical terms, the rule states that utility is maximized when:

$$MUx/Px = MUy/Py \qquad (5.1)$$

where MUx represents the marginal utility received from the last unit of medical care purchased; x, and MUy equal the marginal utility derived from the last unit of all other goods y.

The latter good is often referred to as a composite good in economics. To illustrate why the utility maximization rule should be followed, let us assume that:

$$MUx/Px > MUy/Py. \quad (5.2)$$

In this case, the last naira spent on medical care generates more additional utility than the last naira spent on all other goods. The consumer can increase total utility by reallocating expenditures and purchasing more units of medical care and fewer units of all other goods. As the consumer purchases more medical services at the expense of all other goods (remember that the consumer's income and composite product price are fixed) the marginal utility of medical care falls and the marginal utility of other goods increases. This, in turn, causes the value of *MUx/Px* to fall and the value of *MUy/Py* to increase. The consumer purchases additional medical services until the equality in equation (5.1) again holds, or the last naira spent on each product generates the same additional satisfaction. At this point, the total utility is maximized and any further changes in spending patterns will negatively affect the total utility. The relationship between the utility and medical care is shown in Fig. 5.1.

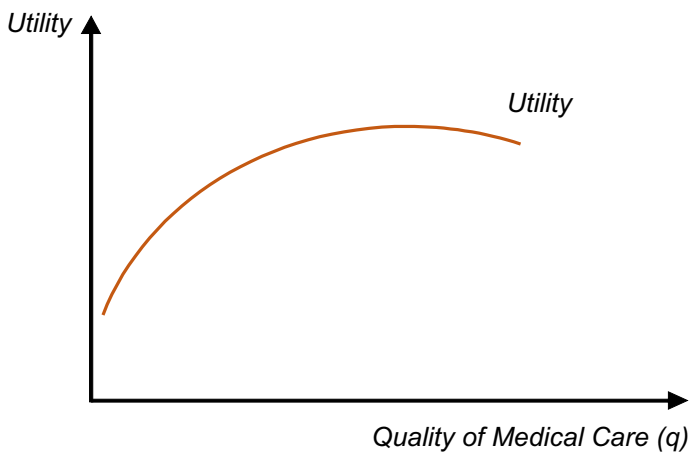


Fig. 5.1. The Relationship between Utility and Medical Care

The Law of Demand

The equilibrium condition specified in equation (5.1) can be used to trace the demand curve for a particular medical service, such as physician services. For simplicity, assume the prices of all other goods and income remain constant, and initially the consumer is purchasing the optimal mix of physician services and all other goods. Now assume the price of physician services increases. In this case, *MUx/Px* is less than *MUy/Py* (where *MUx* and *Px* represent the marginal utility and the price of physician services, respectively). Consequently, the consumer receives more satisfaction per naira from consuming all other goods. In reaction to the price increase, the consumer purchases fewer units of physician services and more units of all other goods. This reallocation continues until *MUx/Px* increases, while *MUy/Py* decreases, and the equilibrium condition of equation (5.1) is again in force such that the naira spent on each good generates an equal amount of utility. Thus, an inverse relation exists between the price and the quantity demanded of physician services.

If the price of physician services continually changes, we can determine a number of points representing the relationship between the price and the quantity of physician services demanded. Using this information, we can draw the demand curve, as shown in Fig. 5.2, where the horizontal axis indicates the amount of physician services consumed (as measured by the number of visit), and the vertical axis equals the price of physician services. The curve is downward sloping and reflects the inverse relation between the price and the quantity demanded of physician services under otherwise equal conditions. For example, if the price of physician services equals P_0 the consumer is willing and able to purchase q_0. Notice that if the price falls to P_1, the consumer purchases q_1 amount of physician services. In this case, price represents the physician out-of-pocket expense the consumer incurs when purchasing medical services from a physician. As such, it equals the amount the consumer must pay after the impact of third-party payments has been taken into account. Naturally, if the visit to the physician is not covered by the third party, the actual price of the visit equals the out-of-pocket expense.

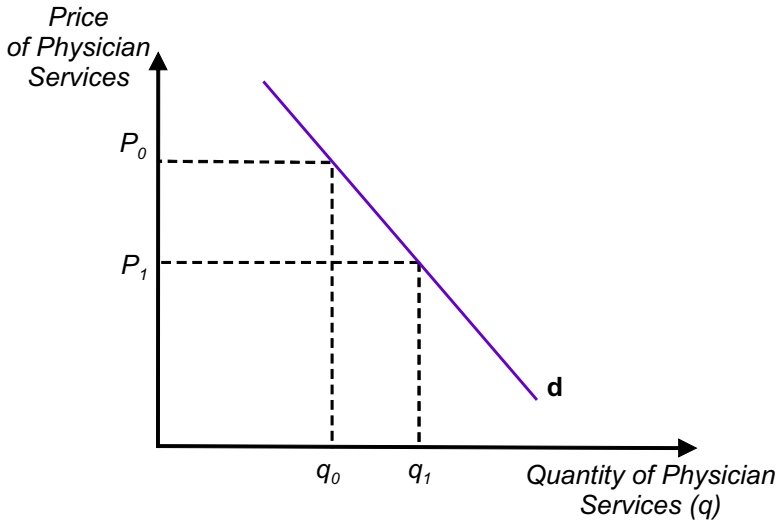


Fig. 5.2. The Individual Demand Curve for Physician Services

The substitution and income effects associated with the price change offer another theoretical justification of the inverse relationship between price and quantity demanded. Both of these effects predict that a higher price will lead to a smaller quantity demanded and, conversely, a lower price will result in a greater quantity demanded. According to the substitution effect, a decrease in the price of physician services causes the consumer to substitute away from the relatively higher-priced medical goods, such as hospital outpatient services, and purchasing more physician services. That is, lower-priced services are substituted for higher-priced ones. As a result, the quantity demanded of physician services increases as the price decreases. According to the income effect, a lower price also increases the real purchasing power of the consumer. Since medical care is assumed to be a normal good (that is, the quantity demanded of medical services increases with income), the quantity demanded of physician services increase with the rise in purchasing power. That also generates an inverse relationship between the price and the quantity demanded because as price falls, the real income increases, and the quantity demanded rises. Taken together, the substitution and income effects indicate that the quantity demanded of physician services decreases as the price increases.

In summary, Fig. 5.2 captures the inverse relationship between the price the consumer pays for medical care (e.g. physician services) and the quantity demanded. The curve represents the amount of medical care the consumer is willing and able to purchase at every price. The utility analysis, or the income and substitution effects, can be used to generate this relationship. This inverse relationship is sometimes referred to as the law of demand. It is important to note that the demand for medical care is a derived demand as it depends on the demand for good health. A visit to a dentist illustrates this point. An individual receives no utility directly from having a cavity filled. Rather, utility is generated from an improvement in dental health. Other economic and non-economic variables also affect the demand for healthcare. Unlike price, which causes a movement along the demand curve, other factors affect the quantity demanded by altering the position of the demand curve.

5.2. The Market Demand for Medical Care. The Fuzzy Demand Curve. Elasticity

The **market demand for medical care**, such as physician services, equals the total demand by all consumers in a given market. In graphical terms, we can construct the market demand curve for medical care services by horizontally summing the individual demand curves. This curve represents the volume of medical services that the entire market is willing and able to purchase at every given price. For example, if the average price of a visit to a doctor is N$50, and at this price consumer A is willing to see a physician three times over the course of a year while consumer B is willing to make four visits, the total, or market demand for physician services is seven visits per year at N$50 per visit. The market demand curve is downward sloping for the same reasons the individual demand curves are downward sloping. In addition, the factors that shift the individual demand curves also shift the overall market demand curve, provided the changes take place on a market wide basis. The market demand curve also shifts if the overall number of consumers in the market increases or decreases. For example, the demand for medical care in a particular community may increase if an influx of new residents occurs. This causes the market demand curve to shift to the right.

The development of a market curve allows us to distinguish between the intensive and extensive margins. The intensive margin refers to how much more or less of the product consumers buy when its price changes. The extensive margin captures how many more or fewer people buy a product when its price changes. Obviously, this is an important distinction to make for a product like medical care. Many medical purchases, such as surgeries, happen only once for a particular individual. As another example, an individual can have a particular tooth pulled only once. This is also a one-shot purchase that either happens or does not happen. If the price of tooth extraction falls, however, we may still observe an inverse relationship between the price and number of teeth extracted. That is, at the extensive margin, more consumers prefer to purchase this one-time form of dental services as price falls. Thus, the quantity demanded may increase with a reduction in price because of changes that occur at the intensive and extensive margins.

The Fuzzy Demand Curve. We have assumed so far that the market demand curve for medical care is a well-defined line, implying a precise relationship between the price and the quantity demanded. In reality, this is usually not the case, and we need to refer to the derivation of the demand curve for medical care to see why. Recall that the demand for medical care is a derived demand and depends on the demand for health and the extent to which medical care influences the production of health.

Important! The relationship between medical care and health, however, is far from exact. That is because there is a considerable lack of medical knowledge concerning the efficacy of certain types of medical interventions. As a result, healthcare providers disagree about the treatment of some certain types of medical problems, and the demand for medical service becomes fuzzy. For example, there is debate among physicians concerning when surgery is necessary for elderly males with prostate cancer. Also, in some instances consumers may lack the information or medical knowledge they need to make informed choices. Consequently, consumers tend to rely heavily on the advice of their physicians when making such decisions as when a particular medical test or surgery is necessary. The implication is that physicians rather than consumers choose medical services, which makes the demand curve fuzzier. Further complicating

matters is the inability to measure medical care produced during a one-hour therapy session with a psychiatrist.

All these factors combined make it extremely difficult to accurately describe the relationship between the price and the quantity of medical care demanded. In other words, the relationship between the price and the quantity demanded is rather fuzzy. A more accurate depiction of the relationship between the price and the quantity may not be a well-defined line, but a gray band similar to the one depicted in Fig. 5.3.

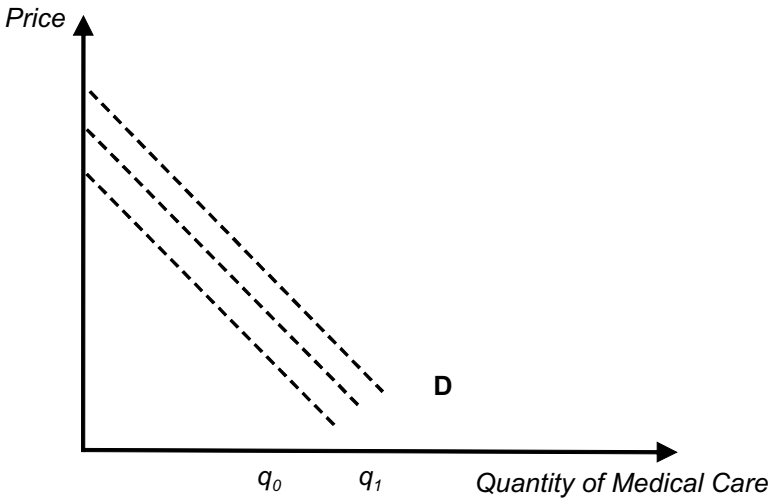


Fig. 5.3. The Fuzzy Demand Curve for Medical Care

There are implications associated with the fuzzy demand curve. First, for a given price we may observe some variation in the quantity or types of medical services rendered. Researchers have documented variations in physician practice styles across geographical areas. Secondly, for a given quantity or type of medical service we are likely to witness price differences. However, the existence of the band is unlikely to detract from the inverse relationship between the price and the quantity demanded for medical care.

Elasticity. The economic theory gives us insight into the factors that affect the demand for medical care along with the direction of their influence. For example, we know that if the price of physician services increases by 15 percent, the quantity demanded

falls. But by how much does it fall? Is there any way to determine whether the decrease is substantial or negligible? The answer is "Yes"; using an indicator that economists call elasticity. Elasticity measures the responsiveness of the quantity demanded to a change in an independent factor.

Own-Price Elasticity of Demand. The most common elasticity is the own-price elasticity of demand. This indicator estimates the extent to which consumers change their consumption of a product or service when its own price changes. The formula for elasticity is:

$$E_D = \%\Delta Q_D / \%\Delta P \qquad (5.3)$$

where E_D denotes the price elasticity of demand;

$\Delta\% \ Q_D$ represents the percentage change in the quantity demanded;

$\%\Delta P$ is the percentage change in price.

From the formula, E_D is a simple ratio that equals the percentage change in the quantity demanded divided by the percentage change in the price. Because elasticity is specified as a ratio of two percentage changes, it is scale free. This makes it much easier to compare elasticity across different goods. For example, we can compare the price elasticity of demand for physician services with that for nursing home care and not have to concern ourselves with the fact that the demand for physician service is usually measured in terms of the number of visits while the demand for nursing home care is measured in terms of the number of inpatient days.

The value of E_D is negative and reflects the inverse relationship between the price and the quantity demanded. In economics, the normal practice is to take the absolute value of the price elasticity of demand measure, or $/E_D/$, and eliminate the minus sign. If the price elasticity of demand is greater than 1 in absolute terms $(/E_D/>1)$, the demand for the product is referred to as price elastic. In arithmetic terms, $/E_D/ > 1$ if the absolute value of the percentage change in the price is smaller than the absolute value of the change in the quantity demanded, or $/ \%\Delta P/ < / \%\Delta Q_D/$. For example, if the price elasticity of demand for dental services equals 1.2, this means the quantity consumed falls by 12 percent if the price of dental care increases by 10 percent under otherwise equal conditions.

The price elasticity of demand is referred to as inelastic if $/E_D/ < 1$, but greater than zero. In this case, $/\%\Delta P/ < /\%\Delta Q_D/$, or the percentage change in the price is greater than the percentage in the quantity demanded in absolute value terms. For example, if the elasticity of demand for physician services equals 0.6, a 10 percent decrease in the price leads to a 6 percent increase in the quantity demanded. If $/E_D/$ equal to 1 because $/ \%\Delta P/$ equals $/ \%\Delta Q_D/$, the price elasticity of demand is unit elastic. This implies that a 10 percent decrease in the price of the product leads to a 10 percent increase in the quantity demanded.

A demand curve that is vertical is said to be perfectly inelastic since no change occurs in the quantity demanded when the price changes. In mathematical terms, E_D equals zero as $\%\Delta Q_D$ equals zero. At the other extreme, if the demand curve is horizontal, it is referred to as being perfectly elastic, and $/E_D/$ equals infinity (∞). Any change in the price leads to an infinite change in the quantity demanded. It stands to reason that the more elastic the demand for the product, the greater the response of the quantity to a given change in the price. Compare the effects of a 10 percent decrease in price of two goods – one with the price elasticity of -0.1, and another with the price elasticity of -2.6. In the first case, the quantity demanded increases by only 1 percent, while in the second case, it increases by 26 percent. We can also use the elasticity of demand to make inferences regarding the slope of the demand curve. Generally, the more elastic the demand for a product, the flatter the demand curve at any given price. This also means that the curve is relatively steep at any given point for an inelastic demand. Consider the two linear demand curves that intersect at point P_0, Q_0 in Fig. 5.4. If the price of the product increases to P_1, the quantity demanded decreases to Q_a of the flat curve.

The own-price elasticity of demand varies across products, and economists point to several factors that determine its value. Among the factors often mentioned are the portion of the consumer's budget allocated to the product, the amount of time involved in the purchasing decision, the extent to which the product is a necessity and the availability of substitutes. As the portion of the consumer's budget allocated to the product increases, the consumer may become more sensitive to the price change. Demand,

therefore, becomes more elastic. An increase in the decision-making time frame may also make demand more elastic. If the consumer has more time to make informed choices, he or she is likely to react more strongly to price changes. Since the consumer typically pays a small portion of the cost of medical services because of insurance, and since medical services are sometimes of an urgent nature, these two considerations suggest that in many cases, the demand for medical services is inelastic with respect to price.

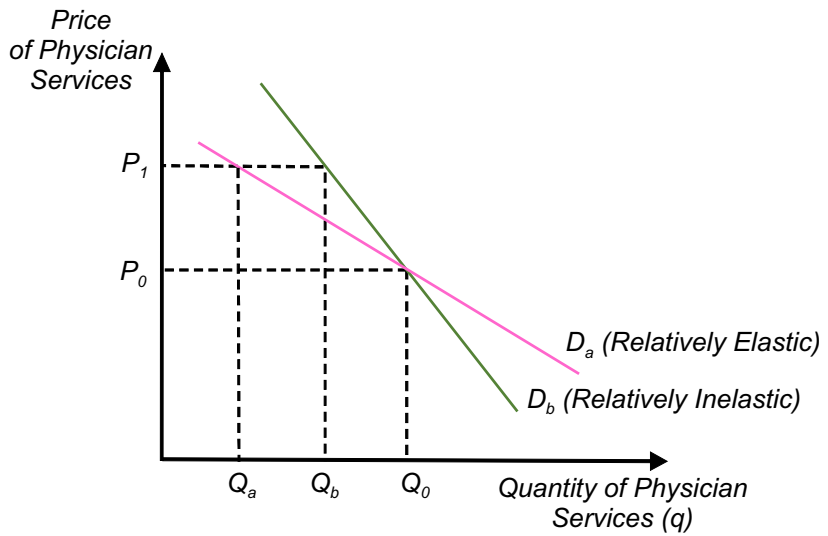


Fig. 5.4. The Elasticity of Demand and the Slope of the Demand Curve

If a product is a necessity, such as a basic foodstuff, the own-price elasticity should be relatively inelastic. The product is purchased with little regard for price because it is needed. Basic phone service might be considered another example of a necessity. Since our society depends so heavily on the phone as a form of communication, it is difficult to imagine an effectively functioning household without it. Naturally, basic healthcare falls into the same category. If an individual needs a particular medical service, such as an operation or a drug, and if its absence strongly affects the quality of life, we can expect that this person's demand will be inelastic relative to the price. In addition, when a person needs a particular medical service in a life-or-death situation, demand is likely to be

perfectly inelastic since the medical service must be purchased regardless of the price if the person has sufficient income.

Considering that many medical services are necessities; we expect the overall demand for medical services to be somewhat inelastic. But this does not mean that the volume of healthcare demanded does not respond to price changes. Rather, it means that the amount of change in the price generates a small percentage change in the quantity of medical services demanded. For some types of medical care, however, demand may be more elastic. Elective medical care, such as cosmetic surgery, may fall into this category since in most instances it is considered a luxury rather than a necessity. As a result, the price may play an important role in the decision to have the surgery. To a lesser degree, dentist services and eyewear might fall into this category. In fact, any medical service that can be postponed is likely to display some degree of the price elasticity.

The availability of substitutes is another determinant of the price elasticity. As mentioned earlier, various types of medical services may serve as substitutes for one another. The larger the number of substitutes, the greater the opportunity to do some comparison shopping. As a result, the quantity demanded of any medical service is likely to be more sensitive to price changes when alternative means of acquiring medical care are available. The own-price elasticity of demand for any given product should be directly related to the number of substitutes available. Stated another way, demand should become more price elastic as the number of substitutes expands. One implication is that the demand for an individual medical care provider is likely to be more elastic than the market demand for medical care. One more point to note concerning elasticity is that the own-price elasticity of demand can be used to predict what happens to total health expenditures if the price increases or decreases. Total revenues (or total expenditures, from the consumer's perspective) equal price times quantity. In mathematical notation,

$$TR = P \times Q_D \tag{5.4}$$

where TR represents the total revenue.

The demand theory tells us that as the price of a product increases, the quantity demanded decreases, or that P and Q_D move

in opposite directions. The total revenue increases or decreases with price changes, depending on the relative rates of change of both variables and the elasticity of demand. Consider an increase in the price of the physician service where demand is inelastic. This means that $/ \%\Delta Q_D / < / \%\Delta P /$, or that the percentage increase in price is larger than the percentage decrease in the quantity demanded in absolute value terms. In terms of equation 5.5, P increases faster than Q_D falls. This means the total revenue must increase with a higher price. If demand happens to be elastic, the opposite occurs: the quantity demanded falls faster than the price increases, and as a result, the total revenue decreases. No change occurs in the total revenue when demand is unit elastic since the increase in price is matched by the same percentage decrease in the quantity demanded.

Other Types of Elasticity. The concept of elasticity can be used to measure the sensitivity of the quantity demanded to other demand side factors as well. The income elasticity of demand represents the percentage change in the quantity demanded divided by the percentage change in income, or $E_Y = \%\Delta Q_D / \%\Delta Y$ where $\%\Delta Y$ equals the percentage change in income. It quantifies the extent to which the demand for the product changes when the real income changes. If EY is positive, the product is referred to as a normal product since any increase in income leads to an increase in the quantity demanded. For example, if E_Y equals 0.78, this means a 10 percent increase in income causes the quantity consumed to increase by 7.8 percent. An inferior product is one for which E_Y is negative, and an increase in income leads to a decrease in the quantity consumed. For most types of medical care, the income elasticity of demand should be larger than zero.

The cross-price elasticity (E_C) measures the extent to which the demand for the product changes when the price of another good is changed. In mathematical terms, $E_C = \%\Delta Q_X / \% \Delta P_Z$ where the numerator represents the percentage change in the demand for product X, while the denominator equals the percentage change in the price of product Z. If E_C is negative, we can infer that the two products are complements in consumption. The cross-prices elasticity between the demand for optometric services and the price of eyewear should be negative. If the price of eyewear increases, the demand for optometric service should drop. Two products are

substitutes in consumption when the cross-price elasticity is positive. For example, the cross-price elasticity of the demand for physician services with respect to the price of outpatient service may turn out to be positive. If E_C equals zero, the demand for the product is independent of the price of the other product.

5.2.1. Other Economic Demand-Side Factors

Income is another economic variable that affects the demand for medical services. Since medical care is generally assumed to be a normal product, any increase in income, which represents an increase in purchasing power, should cause the demand for medical services to rise. Fig. 5.5 illustrates what happens to the demand for physician services when income increases. The increase in income causes the demand curve to shift to the right, from d_0 to d_1, since at each price the consumer is willing and able to purchase more physician services. Similarly, for each quantity of medical services, the consumer is willing to pay a higher price. This is attributable to the fact that at least some portion of the increase in income is spent on physician services. Equally, a decrease in income causes the demand curve to shift to the left.

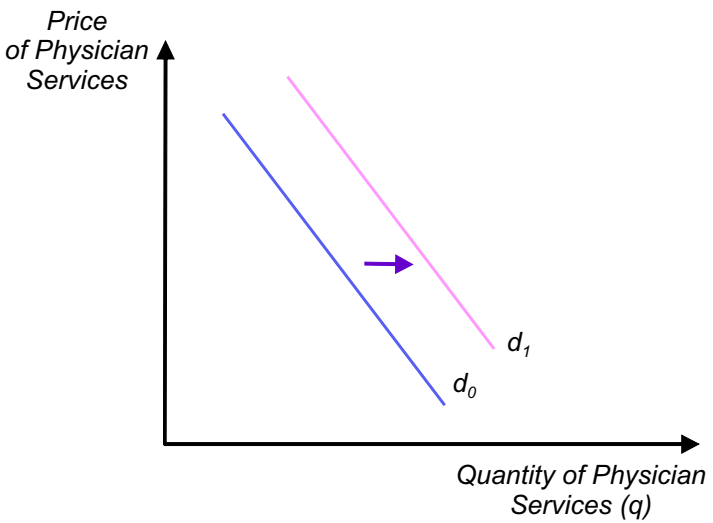


Fig. 5.5. The Shift in the Individual Demand Curve for Physician Services

The demand for a specific type of medical service is also likely to depend on the prices of other goods, particularly other types of medical services. If two or more goods are jointly used for consumption purpose, economists say that they are complements in consumption: as the goods are consumed together, an increase in the price of one product inversely influences the demand for the other. For example, the demand for eyewear (i.e. glasses) and the services of an optometrist are likely to be complementary.

Normally, an individual has an eye examination before purchasing eyewear. If these two products are complements in consumption, the demand for optometric services should increase in response to a drop in the price of eyewear. As a result, the demand curve for optometric services shifts to the right. Another example of a complementary relationship exists between obstetric and pediatric services. An increase in the price of pediatric services should inversely affect the demand for obstetric services. If, for example, a woman postpones pregnancy because of the high cost of pediatric services, her demand for obstetric services also falls. The demand curve for obstetric services shifts to the left.

It is also possible for two or more products to satisfy the same wants or provide the same characteristics. If that is the case, economists say that these goods are substitutes in consumption. The demand for one product is directly related to a change in the price of a substitute good. For example, suppose physician services and hospital outpatient services are substitutes in consumption. As the price of outpatient services increases, the consumer is likely to change consumption patterns and purchase more physician services since the price of a visit to the doctor is cheaper in relative terms. That causes the demand curve for physician services to shift to the right. Generic and brand-name drugs provide another example of two substitute goods. The demand for brand-name drugs should decrease with a decline in the price of generic drugs. If so, the demand curve for brand-name drugs shifts to the left. Finally, eyeglasses and contact lenses are likely to be substituted in consumption.

Time costs also affect the quantity demanded of medical services. Time costs include the monetary cost of travel, such as bus fare or gasoline, plus the opportunity cost of time. The opportunity cost of an individual's time represents the naira value of the activities

the person forgoes when acquiring medical services. For example, if a plumber who earns N$50 an hour takes two hours off from work to visit a dentist, the opportunity cost of the time equals N$100. The implication is that the opportunity cost of time is directly related to a person's wage rate. Taking into account time costs, it is not surprising that children and elderly people often fill doctors' waiting rooms. Time costs can accrue while traveling to and from a medical provider, waiting to see the provider, and experiencing delays in securing an appointment. In other words, travel costs increase the farther an individual has to travel to see a physician, the longer the wait at the doctor's office, and the longer the delay in getting an appointment. It stands to reason that the demand for medical care falls as time costs increase (i.e. as the demand curve shifts to the left).

5.2.2. Non-Economic Determinants of the Demand for Medical Care

The demand for services is influenced by four general non-economic factors: taste and preferences, physical and mental profile, state of health, and quality of care. Taste and preference include personal characteristics, such as marital status, education, and lifestyle, which might affect how people value their healthy time (i.e. their marginal utility of health) or might lead to a greater preference for certain types of medical services. The marital status is likely to impact the demand for healthcare at the market primarily through its effect on the production of healthcare at home. A married individual may demand less medical care, particularly hospital care, because of the availability of a spouse to care for him/her, such as when recovering from an illness.

The impact of education on the demand for medical care is difficult to predict. On the one hand, a consumer with additional education may be more willing to seek medical care to slow down the rate of health depreciation because that consumer may have a better understanding of the potential impact of medical care on health. As an example, an individual with a high level of education may be more inclined to visit a dentist for periodic examination. Thus, we should observe a direct relation between educational attainment

and demand. On the other hand, an individual with a high level of education may make more efficient use of home-produced healthcare services to slow down the rate of health depreciation and, as a result, demand fewer medical care services. For example, such an individual may be more likely to understand the value of preventive medicine (such as proper diet and exercise). In addition, the individual may be more likely to recognize the early warning signs of illness and be more apt to visit a healthcare institution when symptoms first occur. As a result, healthcare problems are addressed early when treatment has a greater probability of success and is less costly. That means that we should observe an inverse relationship between the level of education and demand for medical care, particularly acute care. Finally, lifestyle variables, such as whether the individual smokes cigarettes or drinks alcohol in excessive amounts, affect the health status, and consequently, the amount of the healthcare demanded. For example, a person may try to compensate for the detrimental health impact of smoking by consuming more healthcare services. That translates into an increased demand for medical care. The profile variable considers the impact of such factors as gender, race/ethnicity and age on the demand for medical services. For example, females generally demand more healthcare services than males primarily because of childbearing. In addition, certain diseases, such as cardiovascular disease, osteoporosis, immunologic diseases (such as thyroid disease and rheumatoid arthritis), mental disorders, and Alzheimer's disease, are more prevalent in women than men. Age also plays a vital role in determining the demand for medical care. As stated earlier, as an individual age, the overall stock of health depreciates more rapidly. To compensate for this loss in health, the demand for medical care is likely to increase with age; at least beyond the middle years (the demand curve shifts to the right). Thus, we should observe a direct relationship between age and the demand for medical care.

The state of health controls for the fact that sick people demand more medical services, everything else held constant. The health status and the demand for healthcare are also likely to be directly related to the severity of illness. For example, a person who is born with a medical problem is likely to have a much higher-than-average demand for medical care. In economics jargon, an individual who

is endowed with less health is likely to demand more medical care in an attempt to augment the overall stock of health. Finally, quality of care is also likely to impact the demand for medical care. Since quality cannot be measured directly, it is usually assumed to be positively related to the amount and types of inputs used to produce medical care.

We must also distinguish between a movement along the demand curve and a shift of the curve. A change in the price of medical services generates a change in the quantity demanded, and this is represented by a movement along the demand curve. If any of the other factor changes, such as income or time costs, the demand curves for medical services shifts. This shift is referred to as a change in demand. Thus, a change in the quantity demanded is illustrated by a movement along the demand curve, while a change in demand is illustrated by a shift of the curve.

In summary, the variable we expect to influence an individual's demand for medical care in the economic theory indicates that the demand equation should look something like the following:

$$\begin{aligned} &\textit{Quantity demanded} = f(\textit{out-of-pocket price, income,} \\ &\quad \textit{time costs, prices substitutes and complements, taste} \\ &\textit{and preferences, profile, state of health, and quality of care)} \end{aligned} \tag{5.5}$$

Equation (5.5) states that the quantity demanded of medical services depends on the general factors listed. Note that a change in the first factor results in a movement along a given demand curve, whereas an adjustment in the other factors produce a shift of the demand curve. A rightward shift indicates a greater demand and a leftward shift reveals a lower demand.

5.3. The Supply of Physician Services and Other Medical and Pharmaceutical Services

5.3.1. Inputs into the Production of Healthcare

Physicians. A defining characteristic of the demand side of the healthcare market is the lack of information that individuals have about the cause, nature, and treatment of the disease. Unlike the demand for food, clothing, and other standard consumption commodities, individuals often have poorly defined preferences over

healthcare services. Taking into account this essential feature of demand, physicians of all kinds and specialties play two distinct roles as healthcare providers. First, they provide information and advice to patients on the nature of their condition; the likely impacts of particular treatments, both positive and negative, and their recommended course of action. In addition to these services, physicians engage in the physical delivery services, including surgery, administering of injections, writing of drug prescriptions, and so forth. If individuals were fully aware of the effects of various treatments in improving their health, they would not require the first kind of service (the provision of information), and physicians would be just like bakers and barbers. In these situations, economists find it useful to think of the service provider as the "agent" of the consumer. Of particular reliance is that the agent – the physician or other health care provider in our case – has more extensive information about the consequences and costs of his or her actions than does the patient (who is known in the literature as the "principal").

Two important points regarding this description of roles of physicians should be noted. Firstly, it is possible, at least in principle, to imagine some physicians providing just advisory services, and other physicians providing the executive or operational services, in which case the dual roles would be separated to some degree. This could have important effects on the pattern and cost of services delivered. Secondly, while they are usually better informed than their patients, physicians are unlikely to have full information about the consequences of their actions.

Other Medical Personnel. Physicians make up only a fraction of the personnel resources used in the healthcare sector. Nurses, administrators, clerks, receptionists, traditional healers, and general staff are also included. Some of the labor services provided by such individuals are substitutes for each other and for the work of physicians. For example, trained nurses can administer injections and oral drugs, monitor patients, and so forth. Receptionists could probably handle the temperature measurement, but their performance at the same time is likely to be much lower than that of a nurse because they may make mistakes. Of course, it is not the absolute productivity in performing a task that should determine the allocation of tasks among individuals, but the relative productivity,

or comparative advantage. Thus, even though the physician may be a better typist than the receptionist, it is economically efficient for the receptionist to type up prescriptions and let the physician concentrate on diagnosing and treating patients.

Important! One of the important substitutes for physician services is lower-level medical services (provided by nurses or other clinicians, for example) provided at the appropriate time. It may be possible to substitute the use of a doctor's labor at a time when a patient has developed a severe illness for the use of a clinician's time at a much earlier stage in the patient's life, for example. This is essentially an argument about the possibility of preventive care reducing the need for future curative care. Note that the desirability of this substitution depends on the relative costs of the two types of care, the cost imposed on the patient (in ill health and the like) and other social costs. Most analysts argue that the returns from such a substitution are substantial in sub-Saharan Africa, but some studies in more advanced economies suggest that intensive preventive care, such as screening for certain types of cancer, may or may not be desirable.

Non-labor Inputs. Medical supplies, particularly drugs, instruments, and capital equipment are essential inputs into the production of health services. There may be some degree of substitutability between labor and non-labor inputs, such as consultations with other physicians that might reduce the quantity of drugs required for a given patient, or the use of additional secretariat services in place of office equipment. But many drugs do not have close non-drug substitutes and represent complements to physician care rather than substitute. The absence of these inputs can constrain the productive capacity of the medical services. It is of concern in some developing countries that the quantity of inputs is sometimes very difficult to detect. This is particularly the case with drugs, which often have a non-descript appearance in tablet form. If labeling can be changed at low cost, the scope for fraud is wide and high-quality drugs may be unavailable, except in the black market. This can have two effects: first, treatment may not be provided, and second, treatments with poor-quality drugs may lead to serious unintended effects. In addition, drugs are most effective when prescribed in the correct dosages and in combination with suitable complementary treatments and actions (such as abstinence from drinking alcohol).

While capital goods are important inputs in hospitals, they tend to make up a much smaller share of cost in lower-level outlets, such as clinics. As well as the financial costs involved in procurement of the equipment, resources must be spent to keep them in good working order. These costs are essentially depreciation costs and are incurred with all machines that do not last forever. The problem is that the rate of depreciation can be a function of the uses of the equipment, and the inappropriate use can increase the rate, at which the stock of physical capital becomes ineffective. This, coupled with the possibility that local workers might be untrained in the maintenance of the machines, can impose additional real-resource costs on the healthcare service organization. Other non-labor inputs include the following things.

Drugs. Despite the scope for substitution among inputs, modern drug therapies play an essential role in the treatment of many diseases, and they are second only to personnel costs. Their share of recurrent costs in African economies represents complementary inputs to physicians' visits in health production as is evidenced by large positive correlations between the supply of drugs and the demand for health facility visits. By far the greatest component of the cost of production of most drugs is incurred at the research and development stage, and the marginal costs of production are often close to zero, except when transportation and storage costs are significant. This characteristic of the production process has at least two important implications. First, to give drug companies incentives to incur the large investment costs required, new drugs can be patented, allowing the producer to exercise monopoly power. These patents typically last for a limited amount of time, and when they expire, the drugs can be produced (at low marginal cost) and sold by other companies.

Competition then forces prices down, and drugs become much less profitable. At this stage, drugs with expired patents become known as "generic" drugs. They may have similar properties to other patented (and more expensive) drugs in their effectiveness in combating disease, but they are usually much cheaper.

The existence of products that appear to be different, but are actually similar, means that information problems at the consumer level, which are significant at the market for drugs, might be exacerbated. While one might expect that in such situations uniformed

individuals could choose cheap but inappropriate drugs, it is also possible that they will infer that there are additional benefits associated with the more expensive varieties.

Some accounts suggest that the preference of consumers for specialty drugs over their generic equivalents can be extremely costly.

Hospitals. Hospitals combine a large number of inputs and treat a wide variety of conditions, ranging from the mundane to the exotic. To be able to make sensible policy decisions regarding the allocation of resources to and within hospital, it is necessary to have a model (explicit or otherwise) of how these institutions function and means of explicitly measuring their performance. The standard theory of the firm is not used to us here. On the one hand, the goals and decision-making processes of hospitals are not necessarily well described by the neoclassical model since describing the organization of a hospital in economic terms can be difficult when orthodox notions of ownership and control are ill-defined, and the objectives of those in control are only vaguely characterized. On the other hand, measuring a hospital's performance in productivity terms, for example, is notoriously difficult, given the myriad services they offer.

Pharmaceutical service is all types of services provided by the pharmaceutical personnel as part of the provision of pharmaceutical care. Along with dispensing of medicines, it also includes information, education and advocacy to promote public health, provision of information on medicines and counseling, regulatory activities, professional education and training.

Pharmaceutical services include:

– clinical services aimed at ensuring compliance with the proper use of medicinal products;
– product-oriented services;
– primary healthcare services;
– screening tests in the pharmacy;
– programs to combat drug addiction;
– other advanced services.

Today, the following pharmaceutical services are provided in the world.

1. **Clinical pharmaceutical services.** Control over the proper use of medicines and prevention of side effects are fundamental tasks of modern pharmaceutical practice, which ensure not only optimal treatment of patients, but also cost savings.

Treatment of bronchial asthma

Pharmacists provide a service for patients with bronchial asthma. Their goal is to inform patients about the proper use of inhaled medicines. Such consultations reduce the risk of complications in this group of patients. In 6 countries, the effectiveness of its work has been documented (Belgium, Canada, Finland, Ghana, Norway, and the United States). For example, in Belgium, after 23 months of the program's operation, 36 thousand patients have used the service;

Treatment of diabetes mellitus and hypertension

The activities of pharmacists are aimed at reducing the number of complications in the treatment of diabetes and hypertension. Specialists assist patients in monitoring and controlling the intake of necessary medications. In 4 countries, the results of the programs have been documented (Albania, Canada, Ghana, USA);

Management of anticoagulant therapy

The purpose of the service is to monitor a long-term anticoagulant therapy by determining the international normalized ratio (INR). This is an indicator that determines the rate of blood clotting. This need is often due to the fact that anticoagulants have a narrow therapeutic window. To ensure their maximum therapeutic benefit and reduce toxicity, it is necessary to constantly monitor and adjust the dose.

For example, in France, pharmacists can offer oral anticoagulants to patients subject to the provision of therapeutic support, which includes the following measures:

– in the first year of taking anticoagulants, the patient must attend 2 pharmaceutical consultations, then 1 visit per year. Based on the results of the consultation, the pharmacist assesses the patient's compliance with the prescriptions;

– patients must provide a biochemical blood test to the pharmacist. This is necessary in order to understand how correct the dose taken by the patient is;

– if necessary, the pharmacist can contact the patient's doctor and coordinate the prescription.

During the consultation, the pharmacist informs the patient about the following points:

– how to use the drug;

– the risk of bleeding;

– importance of monitoring blood chemistry parameters;
– medicines that cannot be combined with anticoagulants.

Other clinical pharmaceutical services are:

– drug sales and patient counseling (available in 63 countries);
– the prescription list review (available in 44 countries);
– the tuberculosis program (available in 20 countries);
– the emergency contraception program (available in 51 countries).

2. **Product-oriented services.** These services are related to the following functions of pharmacists:

– preparation of individualized medicines that are not produced on an industrial scale;
– participation of pharmacists in the process of utilization of medicinal products that minimizes harm to the environment;
– control over counterfeit or low-quality medicinal products.

Thus, in addition to controlling the proper use of medicines, pharmacists remain experts in the creation and disposal of medicines.

Product-oriented services include:

– preparation of simple medicinal products;
– preparation of complex medicinal products;
– selection of expired drugs for further safe disposal;
– pharmacovigilance – systematic reporting of adverse drug reactions.

3. **Primary healthcare.** It is the service aimed at informing the population about healthy lifestyles and preventing various diseases (including screening tests).

Primary healthcare includes a vaccination program. The role of pharmacists in immunization and vaccination varies around the world. First of all, they are involved in providing vaccines and promoting immunization. However, in some countries, pharmacists are given the opportunity to play a more active role – to provide vaccinations themselves. Such pharmaceutical activities are authorized in many countries (Argentina, Australia, Philippines, Ireland, UK, USA), often starting with influenza vaccination and gradually expanding. In most cases, such activities require additional training for pharmacists, vaccination scheduling, room and disposal specifications.

Other primary healthcare services are:

– demonstration of injection techniques for various medicines;
– organization of public health campaigns;

– smoking cessation program;
– first aid and follow-up care;
– sampling of biological products for analysis;
– interpretation of laboratory tests;
– participation in the fight against antibiotic resistance.

4. **Conducting screening tests in a pharmacy.** It is a program aimed at conducting screening tests by pharmacists for the early detection of various diseases. This service also allows patients to be referred to appropriate medical centers for further diagnosis. Pharmacists participate in the following screening tests:

– blood glucose control;
– cholesterol level control;
– blood pressure measurement;
– measurement of such parameters as body weight, height, body mass index;
– conducting a pregnancy test;
– HIV test.

Additional screening tests funded by a third party (insurance policies) include:

– France – a test for the diagnosis of influenza;
– Italy, Spain, Switzerland – a test for the colon cancer diagnosis;
– Madagascar, New Zealand – the prothrombin index test;
– Netherlands – a test to determine the kidney function;
– Nigeria – a test for the malaria diagnosis;
– Portugal – the most popular tests for determining the level of triglycerides, uric acid, urea, ALT, ACAT, creatinine, hemoglobin;
– Uruguay – tests for the diagnosis of osteoporosis.

5. **Programs to combat drug addiction.** Programs aimed at reducing individual and social harms associated with drug use include:

– Syringe exchange. The service is aimed at injecting drug users. They are given the opportunity to exchange used needles for sterile ones or replace injecting equipment at little or no cost. The program is available in 19 countries and covers 846 million people.
– Dispensing opioid substitution therapy. The service consists of the introduction of opioid substitution therapy (e.g. methadone, buprenorphine).

6. **Other advanced services and activities are:**
- therapeutic substitution of medications;
- additional pharmacist prescriptions;
- independent prescriptions of pharmacists;
- emergency prescriptions by pharmacists;
- home delivery of medicines;
- provision of pharmaceutical services at night, on weekends and during emergencies.

5.3.2. Incentives and the Allocation of Resources

Considering the technical relationships between inputs and outputs described in general terms in the preceding section, what patterns of resource allocation are we likely to see in the healthcare sector? Such resource allocations can be affected either through direct administrative procedures (for example, by governments) or as a consequence of decisions made by private individuals in the delivery and management of healthcare services. They rely on the choices and behavior of physicians and other providers of labor and physical inputs within institutional settings to implement their preferred resource allocations. This section examines the incentive of agents on the supply side of the healthcare sector, and the implications of different financial and other structures in the determination of resource use.

It is useful to identify two margins on which the allocation of resources may be inefficient. First, incentives may induce agents on the supply side to produce too much or too little care of a particular kind, and second, a given level of care or output may be produced inefficiently with the wrong mix of input, for example, deviation from efficiency on the first of these margins represents allocative inefficiencies, while those on the second represent X-inefficiencies. At a conceptual level, it is not always easy to separate these two notions of efficiency. Taking into account the production function, the allocative efficiency requires that the right quantity be produced, as defined in some way on the basis of the estimated social benefits and costs. However, if production is not undertaken efficiently, the "right" level of output may change, and the allocative efficiency would be defined in some constrained second-best fashion.

Physicians' Objectives and Behavior. It is common in economics to describe resource allocations as the outcome of optimizing behavior of economic agents who have explicit objectives and capacities and who face given financial (and other) reward structures. Within this general framework, physicians can be thought of as both supplying orthodox labor services and acting as entrepreneurs, primarily because of their privileged possession of information about the health production process and the needs of their patients. As the suppliers of labor, their behavior can be analyzed in a similar fashion to that of other workers as resulting from a tradeoff between consumption of goods purchased at the market and leisure. As medical entrepreneurs, the objectives that determine their behavior must be expanded to include such dimensions as effort on the job, prestige and reputation, the well-being of their patients (both generally and specifically with regard to their health), and ethical considerations.

Investing in Human Capital – the Decision to become a Doctor. Faced with market – or government – determined compensation (that is, wage rates), individuals make decisions on whether to become doctors, the type of specialty to undertake, and their hours of work. Like any investment decision, the decision to undertake a typically long-term commitment to study must be made with the opportunity costs in mind. These include forgone current income and any direct financial costs. Of potentially more significance, however, is that, unlike many other investment choices, investment in human capital may be difficult to finance through capital markets. That is, when ownership of individuals is not permitted, a person who invests in human capital through education has no associated physical capital to offer a lender as collateral. Faced with such capital market imperfections, only those with existing assets (that is the well-off) will be able to attain costly education.

5.3.3. Labor Supply

Having received their training, physicians must make decisions about the amount of work they intend to do and where they will do it. The hour's decision is captured conceptually through the standard income-leisure tradeoff as a function of net wages. As net wages

increase, the price of leisure effectively increases, and the substitution effect induces individuals to supply more labor. In opposition to this, higher wages increase each individual's wealth, and the income effect induces the individual to consume more leisure (reduce the labor supply). It is assumed that the first effect dominates, at least over the relevant range, and that the labor supply curve slopes up. Thus, one way to encourage doctors to supply more services is to pay them more.

The location decision presents policymakers with one of the more challenging dilemmas in healthcare provision. Because of the greater availability of both private and public goods and services in urban area, compensation payments to physicians must usually be correspondingly greater in rural areas to persuade them to locate outside major cities. Although money prices of some goods are likely to be greater in the cities, the effective prices of other may well be high in rural areas where the particular item is unavailable. In addition, if incomes are lower in rural areas, the supportable prices that physicians can charge will be lower, so some form of public subsidy is likely to be necessary. In addition to the rural-urban decision, physicians might have the option of locating outside their home country, moving to neighboring countries where their incomes could be higher. This incentive can work in the opposite direction, with doctors trained in western countries sacrificing large salaries to work in developing countries with international agencies. In the presence of significant net out-migration of state-trained physicians, the government may examine the net social return to publicly funded medical education.

Moonlighting is the practice of workers (not necessarily physicians) taking on a second job in addition to their primary employment. Typically, the second job will be in the informal sector of the economy, characterized by a lack of formal administrative structure (labor laws, unions, and so on) and incomplete tax coverage. There are positive and normative issues that may be of concern to a policy analyst with respect to moonlighting. On the one hand, it is useful to understand the forces that lead to multiple job holdings; on the other hand, it is important to know the welfare implications of such activities. Should they be controlled, and if so, how? To understand why physicians might engage in both formal and informal

sector labor supply, we can begin by describing a scenario, in which such multiple job holdings would not be desirable to the individual. Suppose that wages in both sectors are fixed, that the workers can work as much as they wish at the going wage (that is, the labor market is competitive), and that the disutility of work is the same in both sectors. In this case, the optimal labor supply choice is to work in the sector with the highest net wage, and moonlighting would not be observed. However, the conditions imposed in this example suggest situations, in which multiple job holdings would be desirable.

First, if the worker cannot choose hours of work freely, then he may face a constraint in one job that makes the second job attractive. For example, suppose the formal sector wage is higher than that in the informal sector, but no overtime work is permitted (or at least remunerated) in the formal sector, the worker will optimally work as much as possible in the higher-paid job, but may choose to work additional hours in the lower-paid informal sector. Second, the worker may be concerned about the stability of his income. If wages in the informal sector are higher, but more volatile than those in the formal sector, he may wish to provide himself with insurance by working in both sectors. Note that full insurance could be obtained by working only in the low-paid formal sector, but that this may not be optimal if average wages in the informal sector are high enough or if the individual is not too risk-averse.

The third reason to have jobs in both sectors is that there might be private economies of scope in labor supply. That is, working in one sector may increase the individual's productivity in the other sector. The most likely cases are when formal sector employment increases informal sector returns. For example, some kind of training may be provided in the formal sector job that increases both formal and informal sector productivity. Alternatively, attributes of the formal sector job could be used as inputs in the informal sector, such as office equipment and stationery, professional status and reputation, and access to patients (who are treated outside of office hours). Some of these attributes, such as status, human capital, and reputation, have a public good aspect, and their use in the informal sector represents a social economy of scope. Others, however, such as the use of formal sector inputs, permit the worker

to achieve private economies of scope without any social equivalent. That is, the second activity tends to reduce productivity in the formal sector, but it may not reduce the individual's formal sector wage.

The issue of private and social returns to moonlighting suggests that as well as identifying the determinants of this aspect of labor supply, it is important to understand its normative implications. Should all forms of moonlighting be discouraged, and should this be achieved by increasing formal sector wages or increasing penalties on informal sector activities? Should formal sector employment arrangements be made more flexible – either by allowing more discretion in decisions about formal sector labor supply or by allowing outside work – in order to improve the efficiency of labor allocation?

Medical Care Suppliers at the Market. A recurring question in health economics regards the ability of physicians to induce consumers to purchase more medical care than they otherwise would. If such forces exist, it is natural to hope that restricting the supply of doctors might reduce consumption. However, economies usually expect reductions in supply to increase prices. To fully understand these possibilities, it is necessary to examine the interaction of supply and demand.

5.4. Interaction of Demand and Supply-Standard Analysis

So far, we have examined the nature of the production processes of medical care and the incentives suppliers face in determining the type and quantity of care provided and the inputs used. In particular, the influence of prices and other financial variables on supplier behavior was considered. Abstracting from issues regarding the choice of inputs (that is, production techniques), it seems reasonable to suggest, as we did, that higher prices for physicians' services will lead to an increase in the quantity that physicians desire to supply. Although there will be offsetting income and substitution effects, we usually assume that the labor supply curve S_0 (measuring, for example, hours worked) is upward-sloping. In a free, competitive market, wage rate is then determined by the intersection of the supply curve and the demand curve D_0, as in Fig. 5.6. While the

relation between the price received by physicians and the quantity they desire to provide, as described by the supply curve, is likely to be upward-sloping. We usually expect the relationship between the quantity actually supplied and the equilibrium price to be negative for a given demand curve. For example, suppose that the number of physicians in the marketplace increases since the government increases the number of places available at a public medical school or relaxes immigration controls for foreign doctors). There is no effect on the individual supply curves of physicians who were initially active at the market. However, the total supply curve is now the horizontal sum of a larger number of individual supply curve because of the new entrants, and it shift out of S_1, as shown in Fig. 5.6, the equilibrium price of health services has fallen to P_1.

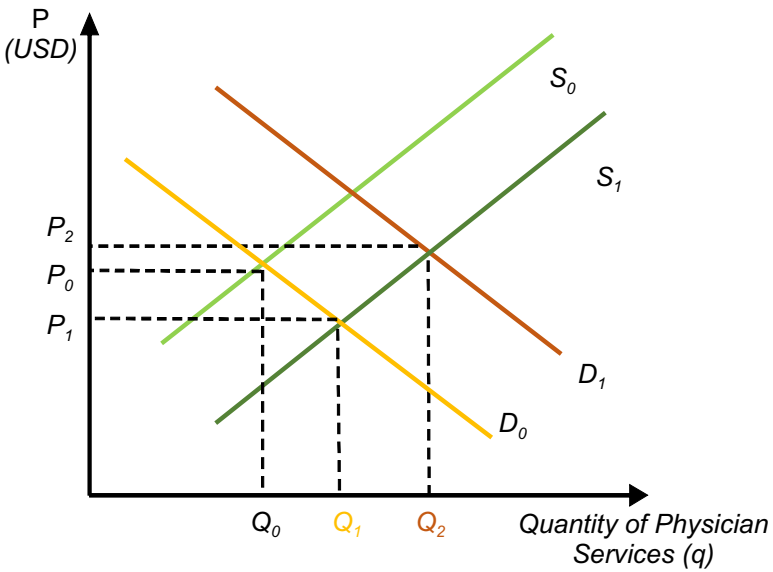


Fig. 5.6. The Demand – Supply Analysis

The market analysis above assumed that the demand curve did not change when the supply curve shifted. This is a very orthodox assumption in the microeconomic theory, the demand curve being the derived relationship between the price and the desired consumption, taking preferences (including health status), prices, and incomes as fixed. The behavior of suppliers, in particular, influences

the equilibrium price and quantity consumed, but not the demand curve. In the case of healthcare, consumers rely on physicians for at least the execution of care. The information provided can relate to the effects of certain treatments, their likelihood of success, likely side effects and other risks associated with them, and the effect of having no treatment. Taking into account this information, individuals make informed choices about the type and quantity of care they desire.

The trouble is, of course, that individuals have little opportunity to verify the information provided, and when the providers of the information stand to gain from providing misleading information, it is unlikely that correct information will be forthcoming. Thus, if the provider of information is the same person as the provider of medical services, individuals may be induced to consume more than they would if they had access to purely objective information. The position of the demand curve might thus be affected by the supplier of services.

This possibility does not answer our question about the ways, in which a positive relationship between equilibrium prices and supply may arise. Indeed, if physicians can influence demand, why do they not push the demand curve out indefinitely and earn higher incomes? There must be some force that restrains suppliers from acting in such a fashion. One possibility is that physicians feel guilt from effectively "fooling" patients into having more treatment than is necessary, and this disutility acts as a constraint on the extent of induced demand. An alternative, and perhaps more natural, constraint on the ability of physicians to induce demand is competition. Over-servicing imposes some costs on individuals (even if they are fully insured against the financial costs of care), and if one physician is found to continually over-service, that individual will lose patients to other physicians. In a perfectly competitive market, this mechanism would restrain the ability of physicians to over-service completely, and they would provide correct information. It is generally acknowledged, however, that the market for physician services is better described as monopolistically competitive because of switching costs. Thus, each supplier exercises a degree of monopoly power over his or her own patients, who incur additional costs if they switch to other doctors. These costs include the time it

takes to find a new doctor with whom the patient feels comfortable, the uncertainty about the quality of the new doctor, and the additional visits, if any, that are required for the new doctor to establish the patient's medical history and condition. Taking into account such local monopoly power, physicians will be able to exercise a degree of demand inducement. The positive relationship observed between the equilibrium price and quantity can then be rationalized by assuming that when faced with an increase in the number of physicians, and thus an outward shift in the supply curve, each physician increases the amount of demand inducement he or she exercises. The effect is to shift the demand curve out to D_1, Fig. 5.7, at the new equilibrium – the intersection of D_1 and S_1 – the price is P_2, and the total quantity is Q_2.

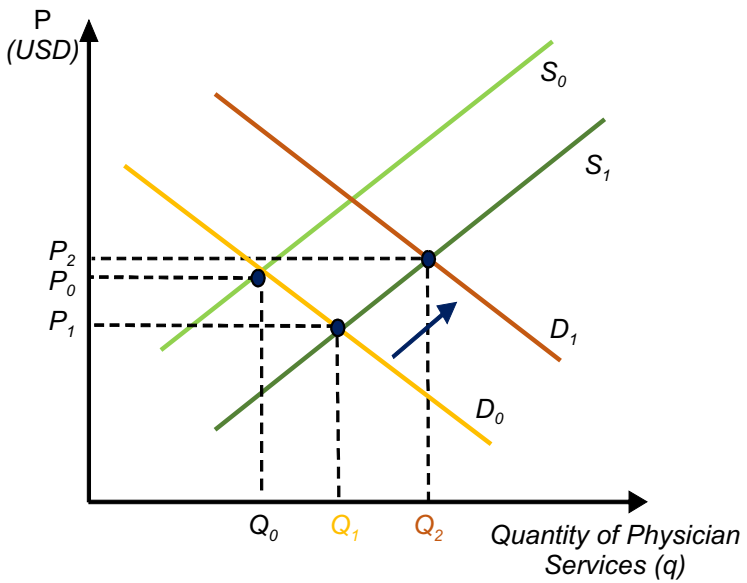


Fig. 5.7. The Supplier-Induced Demand Model

The second mechanism that might increase demand inducement following an increase in supply relates to the extent of the loss a physician experiences when a patient leaves the practice because of over-servicing. The initial increase in supply leads to a fall in unit price. Under our assumption of relatively inelastic demand, this will lead to an increase in the demand by each patient, but to

a reduction in the amount of the expenditures. Therefore, the revenue that the physician earns from each patient falls, and the opportunity cost of losing a patient because of over-servicing also falls. That is, the "price" of over-servicing falls, so we expect more of it. Such a partial equilibrium argument only holds if incomes and other variables are fixed, and we know that the physician's income has fallen because of the reduction in the number of patients, as well as the reduction in revenue per patient. However, the price or substitution effect points in the direction of increased inducement following an outward shift in the supply curve. From the equilibrium discussion above, it should not be a surprise that health economists are divided about the existence of, and underlying mechanisms behind induced demand. There are many studies that attempt to document the existence of the phenomenon, with mixed success.

UNIT 6

HEALTHCARE FINANCING, ITS TASKS AND FEATURES IN A MARKET ECONOMY

Content

Introduction

The precise definition of what services and activities comprise of the healthcare sector is necessary to guide data collection and to make comparisons of health systems across countries or at different times. The following pairs of items show the difficulty of drawing a line between aspects of the healthcare sector/non-health sector. Which should be included within the definition of the healthcare sector, heath services, environmental services (e.g. water, sanitation, environmental pollution control, occupation safety, etc.), hospitals, social welfare institutions, education and training, pure medical research, medical social work, social work, formally trained medical practitioners, traditional medical practitioners? In practice, the boundaries of the healthcare sector vary considerably between countries, and different definitions have been developed for different purposes. In developing countries, the definition tends to be broader than in developed countries due to greater deficiencies in certain areas (e.g. environmental health) and extensive use of the traditional

healthcare sector. A useful rule of thumb is to include all finance/ expenditure which primary intention (regardless of the effect) is to improve health. Financing refers to raising revenue to pay for a good or service. It is the function of a health system concerned with the mobilization, accumulation and allocation of money to cover the health needs of the population, individually and collectively, in the health systems. The whole processes of healthcare finance involve where the money came from, how it was collected and used to pay the providers for their services. This unit discusses healthcare financing.

Objectives

By the end of this unit, students should be able to:

✓ dentify the factors that affect the choice of a financing system;

✓ explore the different sources of financing in the healthcare sector;

✓ understand strong and weak points of different financing mechanisms.

Main Content

6.1. Forms of Healthcare Financing

Financing refers to the ways, in which money is raised to fund health activities, as well as how it is raised to achieve a nation's health objectives. Health financing is a collection of funds from various sources (e.g., government, households, businesses and donors), pooling them to share financial risks across large population groups and using them to pay for services from public and private healthcare providers. Five methods of financing health activities are general and earmarked taxes, social and private insurance, community financing and out-of-pocket payments. However, a financing strategy, which determines how these different methods are combined, is based on the amount of funds available for healthcare, which controls the resources and bears the financial burden. The strategy chosen has implications for the health status and financial risk protection of various income and age groups. Sustainable healthcare systems are built on reliable access to human, capital and consumable

resources. Securing these inputs requires financial resources to pay for investment in buildings and equipment, to compensate the health service staff for their time and to pay for drugs and other consumables. How these financial resources are generated and managed – the process of collecting revenue and pooling funds – raises important issues for policy-makers and planners faced with the challenge of designing systems of funding that meet specific objectives related to social policy, politics and economics. Most countries feel constant pressure since expenditure is increasing, and resources are scarce. According to the WHO, nations have to consider several factors in their selection of healthcare financing methods (Fig. 6.1).

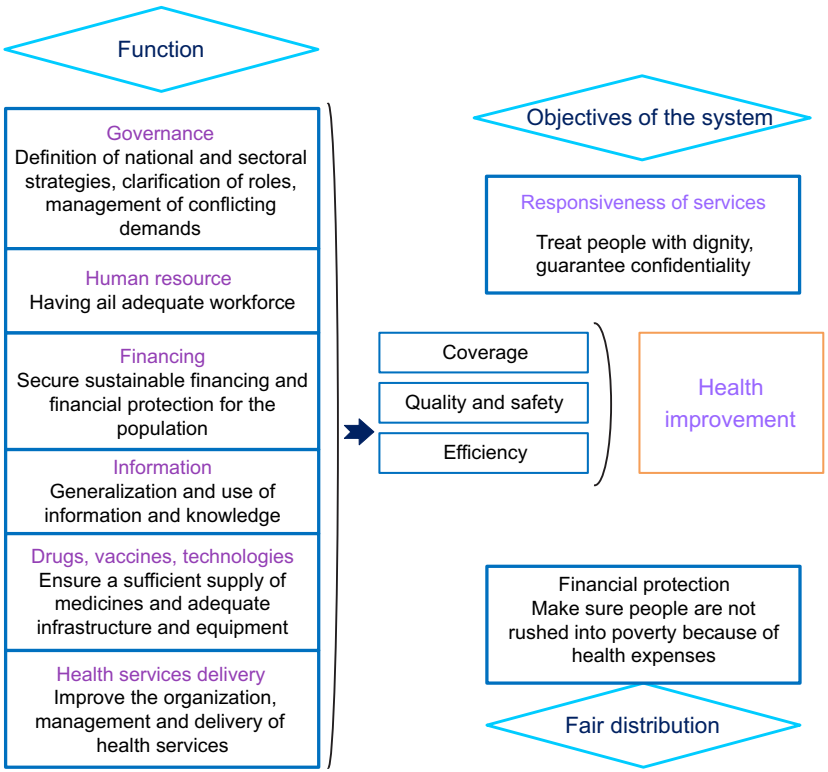


Fig. 6.1. Functions of healthcare systems

These include their fiscal capacity, equity, efficiency in raising funds, and the economic effects of raising the fund. The capacity depends on the context – the fiscal capacity of any method will

depend on the economic structure of the society (the proportion of workers in the formal sector, and on the government's administrative capacity to collect taxes or social insurance contributions). Therefore, when nations search for financing strategies to improve the performance of their health systems, they need to know the relative strengths and weaknesses of the five financing methods. The healthcare system can be broken down into functional components, as shown in Fig. 6.1: revenue collection, fund pooling and the purchasing and provision of healthcare. Functions can be integrated and separated in various combinations. In some cases, the functions are integrated within a single organizational entity; in others, one entity may collect and pool the funds, while other bodies purchase and provide the services. Resources are then allocated between these different entities.

The process of ***revenue collection*** is specifically concerned with who pays, the type of payment made and who collects it. Fig. 6.2 illustrates the diversity of sources of funding, contribution mechanisms and collection agents and how these interrelate. Funds derive primarily from the population (individuals and corporate entities).

Important! The funding mechanisms include taxation, social insurance contributions, private insurance premia, individual savings, out-of-pocket payments and loans, grants and donations. Collection agents can be private for-profit, private not-for-profit or public. Taxes can be levied on individuals, households and firms (direct taxes) or on transactions and commodities (indirect taxes).

Direct and indirect taxes can be levied at the national, regional or local levels. Indirect taxes can be general, such as a value-added tax, or applied to specific goods, such as an excise tax. Some social or compulsory insurance contributions are, in fact, a payroll tax collected by government. Taxes can be general or hypothecated, that is, earmarked for a specific area of expenditure. Social health insurance contributions are usually related to income and shared between the employees and employers. Contributions may also be collected from self-employed people, for whom contributions are calculated based on declarations of income or profit. Contributions on behalf of elderly, unemployed or disabled people may be collected from designated pension, unemployment or sickness funds, respectively, or paid for from taxes.

Private health insurance premia are paid by an individual, shared between the employees and the employer or paid wholly by the employer. Premia can be as follows: individually risk rated, based on an assessment of the probability of an individual requiring healthcare; community rated, based on an estimate of the risks across a geographically defined population; or group rated, based on an estimate of the risks across all employees in a single firm. Government may subsidize the cost of private health insurance using tax credits or tax relief.

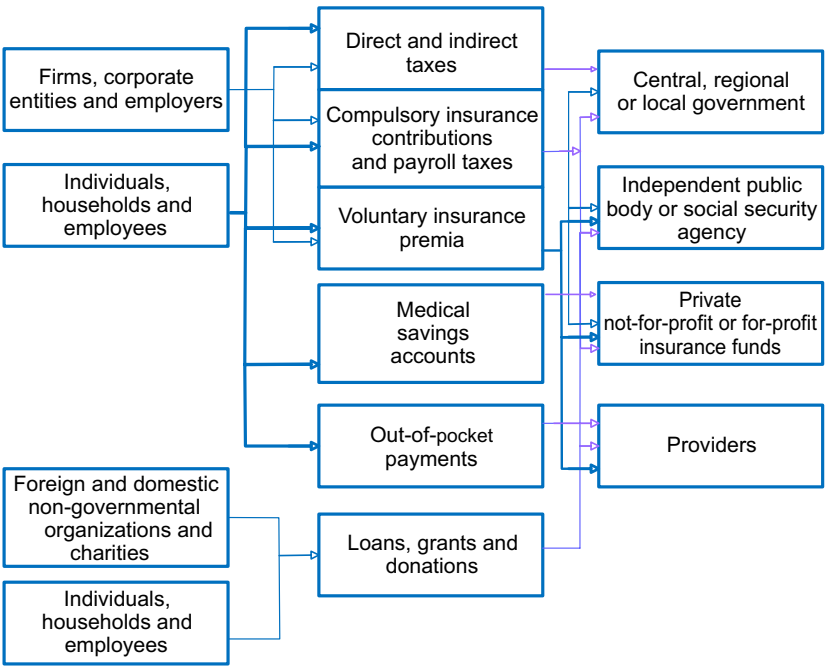


Fig. 6.2. Examples of Funding Sources, Contribution Mechanisms and Collection Agents

Medical savings accounts are individual savings accounts to which people are either required to, or given incentives to deposit money. The money must be spent on personal medical expenses. Medical savings accounts are usually combined with high-deductible catastrophic health insurance. Patients may be required to pay part or all of the costs of some types of care in the form of user charges. These charges may be levied as a co-payment (a flat-rate payment

for each service), co-insurance (a percentage of the total cost of the service) or a deductible (a ceiling up to which the patient is liable after which the insurer covers the residual cost). The collection agent is the provider, such as a physician, hospital or pharmacist.

Fund pooling. Revenue collection must be distinguished from fund pooling, as some forms of revenue collection do not enable financial risks to be shared between contributors, such as medical savings accounts and out-of-pocket payments. Fund pooling is the "accumulation of prepaid healthcare revenues on behalf of the population".

Important! The importance of fund pooling is that it facilitates the pooling of financial risk across the population or a defined subgroup. Examples of this include social health insurance contributions collected by funds and retained by them and national, regional or local taxes that are collected and retained. If different agents carry out these functions, a mechanism is required to distribute resources from the collection agent to the pool. If there are multiple pools, allocation is increasingly being adjusted according to the risk profile of the population covered by each pool. This process is referred to as "risk adjustment". Risk adjustment in competitive social health insurance systems has developed mainly from a concern to prevent cream-skimming. Within tax-financed systems, risk-adjusted capitation methods developed from a concern to ensure equity of access by ensuring a fair allocation of resources to territorial health authorities based on the needs of the population. Irrespective of the source of funds, the underlying rationale for allocating based on risk-adjusted capitation is the same – to ensure that each pool has the "correct" relative level of resources for the population for which it is responsible. Under private health insurance, funds are pooled between subscribers of the same insurance provider. The extent of risk pooling is limited with actuarial premia related to an individual's risk. If premia are community rated, pooling is between high-risk and low-risk members in the same geographic area.

Medical savings accounts prevent pooling by keeping funds in individual accounts. Medical savings accounts are supplemented with catastrophic insurance for expensive treatments. User charges are paid at the point of service and are not a form of pooling. The revenue

generated by user charges is handled differently depending on how the system is designed. For example, the individual healthcare provider may retain the money as income. It may be retained at the level of a clinic or hospital and, together with other revenue, contribute to the cost of maintaining local service provision. If the user charges are given to, or levied by, the insurer or government, they may be used to meet any gap between premium or tax revenue and expenditure.

Purchasing means "the transfer of pooled resource to service providers on behalf of the population for which the funds were pooled". In some systems, separate agents purchase services; in this case, the resources have to be allocated to the purchasers. Pursuing widely held objectives of equity and efficiency requires allocating resources according to the healthcare need. However, many healthcare systems continue to allocate resources based on political negotiation, historical precedent or the lowest bids.

6.2. Sources of National Healthcare Financing Systems

Healthcare financing is a broad term used to define alternative arrangements for paying, allocating, organizing and managing health resources. It includes: defining a level/quality of care preferably the minimum basic health services packages to be provided, in an accessible and equitable manner; identifying different modalities of financing to establish a financially sustainable system; and institute different mechanisms for mobilizing funds and rationalizing the use of available resources, including cost and risk sharing mechanisms/insurances plans.

Financing Strategies The financing mechanisms are grouped into broad and complementary strategies. It includes improving government healthcare sectors efficiency, generating additional and new sources of revenue, encouraging participation of private and non-governmental organizations, development of social and private health insurance, promotion of community participation, encouraging participation of bilateral and multilateral agencies, alternate financing options for the urban areas and organizational mechanisms for implementation of the healthcare and financing strategies.

National healthcare financing systems has pluralistic nature in funding. Therefore, it has different sources of healthcare funding which include

1. Public sources:
– direct government budgeting
– national health services and public services health systems
– social health insurances sponsored or mandated by the government
– community financing
2. Private sources:
– direct payment by households
– private voluntary health insurance
– employers based health insurances
– payments by community and other local organizations
3. External financing; foreign aid or development loans.

Government Financing. Public and QUASI-public sources of Finance:

a) **General tax revenues.** General tax revenue is used in almost every country of the world to finance certain components of healthcare; and in developing countries; it is often the most important source of financing. However, low tax ratios (the proportion of national income collected as tax) in these countries mean that it is often insufficient by itself to support healthcare. Although tax ratios tend to increase in line with development, this depends in larger part on a country's political will to increase the tax burden. In developing countries, general tax revenue is composed largely of duties on imports and exports and sales taxes.

Important! Taxes on business transactions, profits and incomes are all of lesser importance. General tax revenue is currently not the most reliable source of finance for the healthcare sector in developing countries. This results from such factors as the low political priority frequently given to the healthcare sector in national budget decisions; the instability of government finance in countries heavily dependent upon taxes on imports and exports; the frequent use of public expenditure as a tool of the macroeconomic policy; and frequent disparities between budgeted funds and their actual availability or disbursement. The net yield is usually high unless bureaucratic overheads are high. The equity impact of tax systems

is dependent on both the proportional burden of taxation and on the use which is made of the revenue raised. Tax systems can be progressive, falling more heavily on the rich than the poor and, therefore, equitable; but they may also be regressive falling more heavily on the poor than the rich, and inequitable. Developing countries are assumed to have regressive financing systems since they tend to rely on indirect taxation. But, in practice, their tax systems may be progressive since the poorest sections of society fall outside the formal economy and indirect taxes may be levied primarily on luxury items consumed predominantly by the wealthier population groups. Available evidence on the burden of taxation is inadequate to permit often used inequitably in health systems. Health systems are comminuted by high-technology urban-based care and so the rural populations (and the urban poor) have inadequate access to any form of care. There is a limit to what can be collected in tax revenue and how much can be allocated to the healthcare sector without conflict with wider primary healthcare objectives. Taxes that make the poor poorer could seriously damage their health status and undermine their productivity.

b) **Deficit Financing.** General tax revenue may be supplemented by deficit financing that is the decision to borrow and spend funds in the present and repay them over some period of time. Deficit finance may be raised nationally or internationally through such mechanisms as the issuing of certificates or long-term low-interest loans. The cost enjoying the use of those funds in the present rather than the future is the interest that needs to be paid on the loan. In developing countries high inflation rates (affecting the real interest on loans) and the lack of confidence in the government abilities to honor eventual redemption of the bonds may make it difficult to use deficit financing as a source of support for health systems. When it is used, deficit financing is typically for specific construction projects (e.g. hospital water and sewage systems). Unless such projects deliver well, their services or contribute directly to increased output that can be taxed to service the debt, the deficit must be repaid from general tax revenue. Thus, the agency doing the deficit financing must be endowed with the authority to impose additional taxes or fees, or be given a claim on general tax revenue in order to service the debt. Deficit finance may also be raised from

abroad in the form of bilateral or multilateral AID loans, typically given for a project life of between three and five years, and thereby constituting a short-term source of support. Although useful for many developing countries in helping to develop and expand healthcare infrastructure, foreign aid is often limited to support import components.

c) **Earmarked Taxes.** Most tax revenues are paid into a national pool and then shared out between different areas of the government expenditure. Some governments, however, may "earmark" a particular tax for a particular purpose. For example, taxes on the sale of particular products may be earmarked for health services at either national or local level. The problem with such taxes is that they are often difficult to administer, may be politically unpopular and are also often unpopular with tax administrators since they limit their freedom of action. They can be regressive if, as often the cases, taxes are levied on items, such as beer, cigarettes, recreational events, or foodstuffs; but they can be progressive if they are imposed on luxury, products purchased primarily by the more influential sections of society. A clear advantage of this source of finance is that a tax is visibly assigned to priority funding of certain activities or programs. Although not a major source of healthcare sector finance, they may constitute an important source of finance for specific projects.

d) **Social Insurance.** Social insurance can finance healthcare, as well as other needs, such as invalidity and old age support. It is conventionally financed by imposing mandatory insurance payments on employed workers as a percentage of their wages, and by imposing a similar higher payroll tax on their employers. In order to include those workers outside the modern employment sector, insurance payments can also be calculated based on measured income or wealth other than wages, such as the value of crops produced. Allowance will then have to be made for the fact that cash income is only available seasonally, when crops are sold. In their capacity as employers, governments may either run their own social insurance scheme or contract such schemes from private insurance companies. The total financial contribution to social insurance schemes is (in theory) determined actuarially on the basis of the incidence of illness, the conditions of eligibility for benefit,

and the value of those benefits. Individual contributions are not determined, however, on the basis of expected risks or claims, but in some proportion to income. As risks are pooled, there is an unequal benefit distribution in favor of high-risk (high-need) workers. The main problems of social insurance are related to issues of equity and efficiency. It is easiest to cover those in regular employment, who may be as little as 5 to 15 % of the population in developing countries; and there are often marked inequalities in the quantity and quality of services available to those covered by insurance relative to those who are not overall, it is argued that social insurance reinforces the mal-distribution of resources between rural and urban areas in developing countries. It provides extra funds for largely urban, employed workers and leaves the large rural population and the informally employed urban population even further handicapped than before its introduction. Critics of social insurance also argue that it undermines both public and private healthcare by competing with these sectors for limited supplies of real medical resources (e.g. personnel). Finally, it tends to promote or reinforce high-cost, hospital-based, doctor-centered, curative care.

e) **Lotteries and Betting**. These may be used as sources of earmarked income for health and social services in developing countries. Often administrated by quasi-public bodies under national or local government regulation, these typically non-profit schemes rarely constitute an important component of overall healthcare sector finance. Largely supported by the incomes of the poor and thereby constituting a form of regressive taxation, they typically have low net yields because of the payment of prizes and high administrative costs. The typical net yield from lotteries is between 10-30 % of gross receipts.

Private financing for healthcare can be direct or indirect

a) **Direct payment.** This is personal payments made directly to a wide range of providers, including private practitioners, traditional healers and private pharmacists. User fees, whether for government-provided or for privately provided health services, are an out-of-pocket payment and are therefore considered here as health finance from a private source. Similarly, charges to contributions or prepayments by members of community financing schemes are also considered as coming from private (non-government) sources.

b) **Indirect payment**. This is payments for healthcare services by employers (e.g. payment by large and privately owned industrial complexes in developing countries or sharing of healthcare costs by employers in industrialized countries) and health financing by other non-government bodies, such as local charity fundraising for health causes.

Health Insurance:

1. Private Health Insurance. Private health insurance differs from social insurance in two main ways. First, private health insurance typically does not include pensions for invalidity or old age. Second, the price (or premium) charged for private health insurance is not based on the pooled risks of a large population, but on personal risk characteristics and the likelihood of illness in the individual or group covered. As a result, premia are likely to vary for different individuals or groups. Schemes may be profit or non-profit making and may be organized for individuals or groups, the letter often benefiting from lower premia (resulting from lower per capita administration costs, as well as a degree of risk-sharing). In many countries the larger employers act as an organizing body for health insurance, and may pay part of the premium as a fringe benefit. However, in order to control the level of utilization of services, individuals are often required to pay for part of the cost of medical care on a direct fee-for-service basis. In countries where demand is sufficiently high, commercial insurance companies may be active.

Private insurance is not subject to the political allocation process and may channel extra funds into the healthcare sector. However, it suffers from problems of two coverage because of its cost and the exclusion of bad risks, or enhancing inequity and promoting the growth of high-technology healthcare, inappropriate to developing countries.

2. Employer-Financed schemes. In some instances, employers may directly finance healthcare for their employees. They may, for instance, pay for private sector health services, employ medical personnel directly, or provide necessary facilities and equipment. Oil companies, mining and mineral industries, and large-scale export-centered agricultural enterprises usually provide for the health needs of their workforce. Benefits are seldom extended to families as employers are primarily concerned with maintaining the productivity

of the work force. In developed countries, the primary focus is on accident prevention and occupational health, and in developing countries, employers may also have a legal obligation to provide first aid or occupational health services (e.g. sugar and coffee plantations in Latin America, tea and rubber estates in Asia and Cocoa farms and mines in Africa). Problems with employer-financed schemes relate to the quality of care provided, the possible fragmentation of services, difficulties enforcing employer liabilities, and the fact that viability depends upon the performance of the employing agency. Nowhere is employer finance a predominant source of support for health although employer schemes are often a precursor to national social insurance schemes.

3. Charity and voluntary contributions. They can take the form of financial support or in-kind donations (e.g. personal services, physical facilities, equipment and supplies), and may originate from business enterprises, wealthy families, religious organizations, or private individuals. Often these resources are channeled through foundations or religious bodies. The problem with this source of finance is often indirect, for example, donors may have different priorities from the recipient nation and may not recognize their most urgent health needs. They prefer to finance visible evidence of their support, such as physical facilities and equipment, and thereby committing the recipient, country or contributions may also take the place, or reduce other sources of finance. For example, contributions may be eligible for tax relief; reducing general tax revenues for use elsewhere (the effects may be minor). Charitable contributions have played an important role in health services provision in the past, and in some African countries and are still major sources of healthcare finance, channeled through religious agencies. The general trend, however, is for governments to support or take over mission health services. Thus, the role of charitable and voluntary contributions is decreasing although it may still be important in times of emergency and can be a useful supplement to other forms of health finance.

4. Community Financing and Self-Help. Primary healthcare initiatives in developing countries stress the importance of national self-reliance and community participation in healthcare delivery. By mobilizing underutilized national and local resources (e.g., organizational skills, manpower and cash) and by developing affordable

and culturally appropriate delivery systems, it is hoped that basic healthcare will become universally accessible. Consequently, some governments and many non-governmental agencies are turning to communities for organization, participation and financial support, and communal self-help is increasingly thought of as an important source of financial support for health services in developing countries. The challenge is to develop new types of local institutions that can coordinate; and for health services in developing countries to systematically utilize the community resources. Self-help can take many forms, such as labor, local insurance support for volunteer health workers, and drug cooperatives.

5. Direct household expenditure. Household income is ultimately the source of most healthcare finance, but direct expenditure constitutes a specific category of financing that should be considered separately. This category includes any payments that a consumer can make directly to healthcare providers, such as fees for services, or prices paid for goods and supplies. Direct household expenditure is not independent of other sources of finance. Government services may charge user fees (often nominal) for certain services. Even with insurance coverage, there is often a requirement for some degree of copayment, which tends to increase the amount that would otherwise have been spent on health. Health insurance benefits, moreover, may have an upper ceiling, with household requirements in excess of this level. The extent to which these payments represent a real ability and willingness to pay for healthcare is, however, unclear. Willingness to pay does not necessarily reflect the ability to pay. Current levels of household expenditure partly result from the existing pattern of the government healthcare provision, and the limited access to free/cheap government healthcare (particularly in rural areas). People may use and buy non-government (e.g. mission, private, traditional) healthcare partly because they have no cheap or good-quality government alternative. Low-income groups tend to delay use of health services until illness is severe, presumably in part to avoid payment, but such delay generally only increases the necessary expenditure. High healthcare bills may sufficiently undermine their economic position to push them further into poverty. Healthcare payments also sometimes displace expenditure for other basic necessities of life (e.g. food) since there is only a limited ability

to pay for the range of household needs. Utilization and payment for health services is, moreover, likely to depend heavily on the perception of their relevance to a specific health need and the extent to which they provide a service that people value. The use of traditional healers, for example, may reflect a belief in the relevance of their treatments for certain diseases rather than a general willingness to pay for any type of healthcare. Perceptions of poor quality in government services certainly undermine their use, and therefore, willingness to pay for them. Private services may be more oriented to the preferences and circumstances of households, for instance, providing for pay. Raising the level of direct household expenditure for healthcare, for example, by user fees, will clearly have a negative impact on equity (by influencing both the distribution of the payment burden and the benefits failed). It may be mitigated by the introduction of an exemption mechanism for the poor although such a mechanism may itself reduce the demand for healthcare made by low-income groups since they may not wish to be identified as "poor". Moreover, such willingness to pay as exists is attached primarily to curative services, and so can only extend the provision of preventive care if it is possible to re-allocate resources within the healthcare sector. Finally, the potential yield from user fees is unclear. It is dependent on the level and type of fees, the bureaucratic structure required to implement them, the existence of exemption mechanisms, and the impact of fee systems on the demand for care and the rates of collection. The administrative difficulties of implementing a fee system (e.g. how is the ability to pay assessed? Who assesses it? Who collects the fees? How is abuse of the system restricted?).

6. Health Insurance. It provides the means by which risks or uncertain events are shared between many people. Premia are paid to an insurance institution, which compensates any insured victim of the event for any financial loss resulting from the event. Insurance, therefore, helps to lessen and spread risks, and it relies on the fact that what is unpredictable for an individual is highly predictable for a large number of individuals. It follows that for insurance to be feasible, there must be enough individuals insured to spread the risks widely, and the uncertain events must be relatively independent of each other. That is, the principle is one of insurance based

on probabilities, not one of prepayment for known future events; though in practice, a prepayment element for healthcare exists since certain types of utilization are highly predictable. For a health insurance scheme to be cost covering, the level of its premia needs to be related to the statistical frequency with which the population covered requires care, and to the average cost of claims, plus an allowance for administrative costs and a profit margin (for commercial institutions).

Important! Insurance has redistributive consequences, their nature and magnitude depending on the financing of the schemes and the way in which premia are assessed. Since the occurrence of the event being insured against is uncertain, some participants will draw out more than they pay in, thus resulting in redistribution from the healthy persons to the sick. Other distributive effects will depend on whether the insurance is organized privately or through collective mechanisms, and on the method of distributing the costs over the population.

Health insurance can be financed and organized in a variety of different ways. It can be purchased by an individual or group through the private market, from either profit or non-profit firms, and under these circumstances is conventionally termed private or voluntary health insurance. Healthcare itself would usually be delivered by independent providers, but sometimes by facilities owned by the insurer. In the case of private or voluntary health insurance, the level of an individual's premium would be based on the actuarially determined likelihood of illness of that individual. In contrast, group insurance is often based on a firm or co-operative and the premia related to the risk of the group of employees in it, not of individuals. All subscribers will pay similar premia and such insurance may well be made compulsory by the firm to prevent low risk or high-income employees opting out. In some countries (for instance the United States and Australia) there are examples of the imposition of community rating on private insurers; that is, within a given geographical area, premia are not permitted to vary according to health risk or occupation. Premia are often paid at least in part by employers, health insurance is considered a fringe benefit, though labor legislation making it compulsory for employers to provide their workers with some form of medical care is increasingly being introduced in developing countries.

6.3. Factors Affecting Healthcare Financing

The form and level of healthcare financing are now major policy issues for most developing countries and it is essential that decision makers have a clear understanding of the implication of alternative approaches to financing healthcare. There is an increasing interest in how health services are funded in all countries. The following factors, among others, influence the health services sector and should be given due attention in healthcare financing.

1. Demographic Changes. These have major effects on healthcare provision; firstly, demographic change may lead to variations in the health coverage of the population. Rapid population growth rates can cause tremendous strains on the provision of social services, including healthcare. Secondly, the age structure of the population has an important significance to the provision of healthcare. There are higher health service unit costs associated with the young and the old. The antenatal, obstetric and under five age groups are all heavy users of healthcare, as are the elderly with their higher incidence rate of chronic illness. Thirdly, demographic factor relates to the relationship between economic producers and dependents of a country. A high dependent ratio means an increased burden on the productive population for providing healthcare.

2. Economic Recession. This can be expressed by low or even negative growth rates, increasing debt burdens and high inflation rates. This has severe implications for the ability of governments to maintain, let alone expand, expenditure on healthcare. Such effects on the supply of healthcare are worsened by the increased need for healthcare brought about by the recession itself through the links between poverty and ill health.

3. Rising Expectation. The rising expectation of healthcare consumers especially the middle classes, to receive high-technology medical care similar to that available in the industrialized world is high.

4. Concerns about equity. Governments committed to the principles of primary healthcare have a major responsibility to improve levels and depths of coverage. The concerns for equity may influence the choice and system of healthcare financing. To extend basic

healthcare at a time when there is such strong middle-class pressure may only be available by providing substantial additional resources to the healthcare sector.

5. Disease-pattern changes. Disease-pattern change may result due to changes in average income levels or due to changes in social development. Thus, as standards of living rise and morbidity patterns change, these changes are likely to have an effect on healthcare financing. In addition to shifts in disease patterns, the advances of medical technology have led to the possibility of treatment for health problems previously accepted as untreatable. This again places further pressures on healthcare providers.

6. Efficiency. Given the limited resources available for health in developing countries, it is essential to taste and use resources as efficiently as possible.

7. Displacement effects. Rather than generating additional resources for the healthcare sector, new or expanded financing mechanisms may merely displace funding from other sources. Displacement is not necessarily an undesirable consequence if the new or expanded source of finance is more efficient or more equitable than the one it partially displaces. Examples of displacement effects include foreign assistance, which may displace government support for healthcare; counter-funding often a precondition for foreign assistance, which may divert funds away from existing priority projects; health insurance schemes, which may in some instances displace earth than additional to the total of resources being allocated to healthcare (e.g. displacing direct payments); charitable contributions, which may be withdrawn when other sources are developed; and government allocations, which may be reduced when other sources of finance (such as user fees) are developed.

8. Wider effects of the healthcare sector. Healthcare sectors may account for a sizeable share of national resources and are often major employers.

Important! Consequently, the activities of the healthcare sector may have spill-over effects on the economy as a whole. These include external effects on costs (e.g. inflation through the repercussions of high increases in stag pay); foreign exchange problems through heavy foreign borrowing for development projects or for payments for imports, such as pharmaceutical, or equipment opportunity costs,

such as the attraction of scarce manpower into the healthcare sector at the expense of other professions, and disincentives to investment and employment (e.g. as a result of financing health services through high taxes on certain economic activities, enterprises or sectors). These external effects may also be positive as in the case or improved productivity resulting from reduced death and disability in the work force.

In selecting a system of financing healthcare some criteria should be used. The first three criteria outlined below are general, while the last two have particular importance within the context of primary healthcare:

a) Viability and ease of using the system. This implies bureaucracy and cost simplicity, social acceptability and technical feasibility.

b) Revenue generating ability. Net revenue minus earning ability = Revenue minus operating costs. The administration of user changes, for example, may include the costs of billing, accounting and the safe storage and collection of funds. Even where the additional staff is not employed, and the existing staff are used, it implies an opportunity cost to the health service in terms of alternative activities that the staff could engage in if they were not involved in the revenue generating scheme.

c) Effects on service provision. Systems of financing, for example, which involve three parties – a patient, a provider, and an insurance company – may lead to over-provision of certain services.

d) Effects on equity. That is equal access to care for those in equal need.

e) Participation in decision-making. This is a concept that stresses community participation, which creates an opportunity for a direct relationship between a consumer and a provider; an example of a financing system suitable of such participation is user charges.

UNIT 7
GOVERNMENT AND ITS ROLE IN HEALTHCARE

Content

Introduction

In recent years, healthcare reform has come to the top of the political agenda around the world. For developed industrial countries and many middle-income developing countries, reasons include rapidly rising costs, the large number of people still not covered by health insurance and the fear of AIDS. For developing countries, the main reason is a better understanding of the importance of health for improving the productivity of workers and the potential for enormous gains in health at very low cost. Governments all over the world have played a vital role in bringing about the great advances in health over the past many years. Public health measures are responsible for eradicating smallpox and have been central to the reduction in deaths caused by other vaccine-preventable childhood diseases. Expanded and improved clinical care by government doctors and nurses has saved millions of lives from infectious diseases and injuries. Better prenatal and delivery services organized by governments have lowered the rate of serious

complications of pregnancy and childbirth for millions of mothers. Despite these remarkable improvements, however, enormous health problems remain. Absolute levels of mortality in developing countries are still unacceptably high; child mortality rates are about ten times higher than those in the established market economies. According to the World Bank Development Report, if death rates among children in poor countries were reduced to those prevailing in the rich countries, 11 million fewer children would die each year. Almost half of those preventable deaths are as a result of diarrheal and respiratory illness exacerbated by malnutrition. In addition, every year seven million adults die of conditions that could be inexpensively prevented or cured; tuberculosis alone causes two million of these deaths. Over 400,000 women die from the direct complications of pregnancy and childbirth. Maternal mortality ratios are on average 30 times as high in developing countries as in high income countries. There are several major problems with the way health systems are now run and financed and if solutions are not found, the pace of progress in reducing the burden of premature mortality and disability will be slowed. The appropriate nature and extent of government involvement will vary from country to country, in part depending on income levels.

Objectives

At the end of this unit, the students will be able to:

✓ understand the role of government as affecting the resource allocation pattern in health and the extent to which it can influence the overall performance of the sector;

✓ analyze the possible measures that can be taken to alleviate the health problems of developing countries;

✓ appreciate the problems of health policy in developing countries.

Main Content

7.1. Problems of Health Policy

Some of the common problems of most countries in their policy are misallocation, inefficiency and cost allocation.

a) Misallocation. One of the most important aspects of economics in making health policy is the appropriate allocation of material,

financial and human resources. This implies optimal distribution of economic resources among competing needs. This calls for the proper identification of the need. Sometimes public money is spent on health interventions with low-cost effectiveness, such as cancers, and critical and highly cost-effective interventions, such as treatment of tuberculosis and sexually transmitted diseases, remain under funded.

b) Inequity. The poor lack access to basic health service and receive low-quality care. Government spending for health goes disproportionately to the affluent in the form of subsidies to sophisticated public tertiary care hospitals and to private hospitals.

c) Inefficiency. Much of the money spent on health is wasted because brand name pharmaceuticals are purchased instead of generic drugs, health workers are badly deployed and supervised and hospital beds are under-utilized.

d) Cost explosion. In some middle-income developing countries, healthcare expenditures are growing much faster than incomes as the increase in the number of specialists, the availability of new medical technologies and the expansion of health insurance linked with fee-for-service payments together generate a rapidly growing demand for expensive tests, procedures and treatments. As countries alike rethink the best way to provide healthcare in the century ahead some argue that governments should step up their financing while allowing more participation by non-government organizations and the private sector in supplying services.

Governments need to be involved in *finding solutions* to the problem of the healthcare sector based on the problems mentioned. The poor cannot always afford the healthcare that would improve their productivity and well-being. Some actions to promote health are pure public goods or care large positive spillover effects. Market failures in health insurance also mean government intervention can raise welfare by improving the functioning of these markets. Clearly, governments have a responsibility to spend wisely and to evaluate carefully exactly what form their involvement should take. The World Bank recommends four main policies to overcome the existing weakness of health systems in developing countries.

Governments should *finance a nationally defined package* of essential public health and clinical care, especially for the poor,

and should ensure the widespread and efficient delivery of such a package.

1. The public sector should devote far fewer resources or none at all, to financing health services outside of the essential package which are of lower cost effectiveness.

2. Governments should promote such types of health insurance that not only achieve broad coverage of the population, but also build in payment mechanisms that control the cost of health services.

3. Governments should encourage diversity and competition in the supply of health inputs, particularly drugs, supplies and equipment, as a means of improving quality and driving down costs. They should also foster a competitive private sector to provide the full range of health services, including financial publicity.

7.2. A Basic Health Package

Government action in many areas of public health has already had an important payoff. The challenge now is to expand coverage of interventions with high cost-effectiveness. School based health services information on family planning and nutrition programs to reduce tobacco and alcohol consumption regulation, information, public investments to improve the household environment and AIDS prevention could be explored. At the same time, governments should also put together a package of essential clinical services, depending on local needs and the level of income. The World Bank Development Report has come out with a suggested minimum package of health services which is affordable by developing countries at their levels of health spending and would reduce the burden of disease by just over 30 percent in low-income countries. Eleven clusters of interventions or individual interventions are included in the package, apart from being cost-effective these services address diseases responsible for a large share of the disease burden in developing countries.

However, the exact content of each country's essential package will be largely determined by the epidemiology profile of the country (the distribution of disease burden across diseases) and the cost effectiveness of the corresponding interventions. The size of the

package (number of intervention cluster) will depend on the financial resources available for healthcare. Clustering interventions improve cost-effectiveness through at least three mechanisms:

1. Synergism between treatments or prevention activities is common, particularly in pediatric care.

2. Joint production costs can substantially reduce the amount of resources needed to carry out interventions separately.

3. The optimal use of specialized resources, such as hospital beds, requires a screening process to refer the most severe cases from the first level of care to other facilities.

An efficient health cluster should include interventions that can be given to the same individual, at the same time, and through the same mode of delivery (outreach community health worker, health center or hospital). The expanded program on immunizations, for example, is a very efficient one as it includes six vaccines provided through the same delivery system to the same individuals, often at the same time, hence, an essential health package approach is an important measure that governments can be encouraged to do.

7.3. Redirecting Public Spending

The World Bank Development Report pointed out the need for widespread and fundamental reform of health policies and health systems. It called for changes in the level and composition of government spending for health in public and private institutions responsible for delivering health services and in insurance, cost recovery and mechanisms for financing healthcare.

Important! Public financing of an essential clinical package can be justified since the package creates positive spillover effects and reduces poverty. However, the case for government financing of discretionary clinical health services outside of the essential national package is far less compelling. In fact, if governments reduced or eliminated public funding of these services they would actually increase in both efficiency and equity. One important way to direct government spending away from discretionary care is to recover costs in government hospitals especially from the wealthy and the insured. Even in low-income countries, such as Nigeria, Kenya, Pakistan and the Philippines, where insurance may account for less

than 5 percent of the total health spending, a combination of limited private insurance and the ability of upper income groups to pay makes feasible for governments to charge for discretionary care delivered in public hospitals. In middle-income countries where insurance becomes more important there is an increasing potential for cost recovery.

Governments should also phase out public subsidies to insurance which generally benefit the better-off. There are strong efficiency arguments for directing government funding to public health interventions because of the public good nature of these services and a number of the essential clinical services, including treatment of the effects of tuberculosis. In addition, there are equity grounds for financing the basic health package. The poor are disproportionately affected by the disease burden the package addresses. This means that making public financing of this package with universal government finance leads to public subsidies to the wealthy, who can afford to pay for their own services, with the result that fewer government resources go to serve the poor. One way to solve this problem is by targeting public spending to the poor. In low-income countries where current public spending for health is less than the cost of the minimum package some targeting is almost inevitable. In countries where the wealthy do not use government financed services because of the greater quality and convenience of privately financed services, targeting may be fairly easy. The most sophisticated facility required to deliver the minimum elements of the package is a "district" hospital, which serve as the first level of referral from health centers. These hospitals offer basic surgery, emergency services and some outpatient care. Generally, they can have 100-400 beds and serve 50,000-200,000 inhabitants; the minimum package requires access to health centers and district hospitals throughout the country. On the average, it requires about 1 district hospital bed, 0.1 to 0.2 physicians per 1,000 population and 2-4 nurses per physician. Governments can direct public spending to support the nationally defined ***essential package*** in various ways.

a) Where services are publicly financed and provided, government can reallocate public spending towards inputs – drugs, supplies and equipment, staff and facilities that support the package. In many countries extending lower-level facilities are necessary steps

to delivering the package. At the same time, governments can eliminate or greatly reduce financing of inputs for less cost-effective services. This might include losing wards or converting specialized hospital physicians.

At the same time, the provider's treatment decisions would not be micromanaged; they would be influenced by the nature of the input availability. The specialized staff and equipment, for example, would be available for treating malaria in young children. Budgetary and salary incentives could also be used to reward individual providers, facilities that achieved good coverage of the population with the services in the package.

b) Where services are publicly financed, but privately provided, governments should reimburse only for those services in the essential package. This model of healthcare delivery is growing. However, in developing countries, it is still uncommon. At present the regulatory capacity to oversee such arrangements is poorly developed.

7.4. Controlling Costs

Where subsidies in discretionary clinical services for the better-off are cut or public insurance is universal and pays for a more comprehensive set of services in the national package, governments must cope with the problem of escalating healthcare costs. These costs can crowd out spending on other sectors of the economy or raise the price of labor threatening a country's international competitiveness. The sources of excess health costs are complex and much debated. Health services are labor-intensive, and their productivity grows slowly compared to the other areas of the economy. In the United States, higher levels of underlying morbidity and greater hospital amenities relative to other industrial countries are part of the answer. But two types of inefficiencies are also important, high administrative costs and unnecessary use of an ever-expanding array of costly technologies of diagnostic tests and surgical procedures. These inefficiencies appear to be linked to two basic features of the US health system. Open-ended free-for all service compensation for health providers encourages the development of new equipment, drugs and procedures since neither providers nor patients have strong incentives to hold down utilization or spending. A complex

system of multiple insurance institutions and other payers, each with its own procedures, raises administrative costs substantially. These findings concerning health costs escalation in industrialized countries are especially relevant for middle-income developing countries, which are under pressure from medical professionals, manufactures and consumers to use new medical techniques. They face difficult policy choices related to provider compensation. One approach to controlling health costs is to pay a fixed amount for each person (capitation) as is done by health maintenance organizations. Another approach used in several industrial countries is to provide each hospital or network of physicians with a fixed total budget. In countries where there is expanded insurance system, insurers can jointly negotiate uniform fees for physicians or they can set fixed payments for specified medical procedures.

7.5. Strengthening Household Capacity

Within the household, health improves as people escape poverty and get better education. Beyond the household, every society's health services are affected by its national income and its ability to acquire and apply new scientific knowledge, which depends on the level of schooling.

The role of income. Life expectancy is believed to be strongly associated with income per capita. The higher the country's income per capita, the more likely its people are to live long, healthy lives. The income growth has more impact in poor populations because additional resources buy basic necessities, particular food and shelter that yield especially large health benefits. The relationship between income and life expectancy has improved over the course of the century as advances in science and medicine have made it increasingly possible to realize greater health for a given income.

Since poverty has a powerful influence on health, it is not just income per capita that is relevant. The distribution of income and the number of people in poverty matter as well. In industrial countries, the life expectancy depends much more on income distribution than on income per capita, and it has been rising faster in countries with the improving income distribution. In developing counties, the variation in the prevalence of poverty and per capital

public spending on health goes a long way toward explaining differences in life expectancy. Moreover, the adverse effect of poverty on health can be seen in health differences across rich and poor neighborhoods and families, even within the same city.

The Role of Education. Households with more education enjoy better health for both adults and children. A mother's schooling is a powerful determinant of child health. The advantages that a mother's schooling confers on her children's health are felt even before birth, and they continue to operate throughout the childhood years. Better-educated mothers marry and start their families later diminishing the health risks of early childbearing. They also tend to practice better domestic hygiene and make more effective use of health services. In general, they are better at getting information on health and acting on it. Among adults, health depends strongly on personal habits and lifestyles. Since educated people tend to make choices that are better for their health, there is a strong relation between schooling and health. In Brazil, adults with primary schooling or less are about five times as prone to high blood pressure as those with post-secondary schooling. Educated people are quick to modify their behavior as new health threats arise (such as AIDS) or in response to new information about health. In the United Kingdom, for example, the proportion of smokers among adults declined by 50 percent between 1958 and 1975 among the most educated, but hardly changed among the least educated.

Taking into account these close links between better health and income and education, the policy implications are clear; governments should work to boost economic growth, reduce poverty and expand schooling (especially for girls – one of the most effective ways of strengthening the women's ability to care for their families). It is difficult to reduce poverty and thereby improve the health status without the economic growth, so establishing sound economic policies is one of the most valuable things a government can do.

REFERENCES

1. Cost-utility analysis of lifestyle interventions to prevent type 2 diabetes in women with prior gestational diabetes / A. Werbrouck [et al.] // Eur J Public Health. – 2020. – Vol. 30, No. 3. – P. 473-478.

2. Currie, J. Diagnosing expertise: Human capital, decision making, and performance among physicians / J. Currie, W. B. Macleod // Journal of Labor Economics. – 2017. – Vol. 35, No. 1. – P. 1-43.

3. Dafny, L. More Insurers Lower Premiums: Evidence from Initial Pricing in the Health Insurance Marketplaces / L. Dafny, J. Gruber, C. Ody // American Journal of Health Economics. – 2015. – Vol. 1, No. 1. – P. 53-81.

4. Davis, M. Selling Immunity Self, Culture and Economy in Healthcare and Medicine / M. Davis. – London : Routledge. 2022. – 160 p.

5. Determinants of Health. Healthy People [Electronic resource]. – Access mode: https://www.healthypeople.gov/2020/about/foundationhealth-measures/Determinants-of-Health. Updated August 23, 2018. (Date of access: 02.10.2023). – The name from the screen.

6. Dewar, D. M. Essentials of health economics / D. M. Dewar. – New York : Jones and Bartlett Publishers, 2017. – 150 p.

7. Dranove, D. Employed for Higher Pay? How Medicare Payment Rules Affect Hospital Employment of Physicians / D. Dranove, Christopher Ody // American Economic Journal: Economic Policy. – 2019. – Vol. 11, No. 4. – P. 249–271.

8. Economic Analysis in Healthcare / Stephen Morris [et al.]. – New York : Wiley, 2012. – 390 p.

9. Geruso, M. Upcoding: evidence from Medicare on squishy risk adjustment / M. Geruso, Timothy Layton // Journal of Political Economy. – 2020. – Vol. 128, No. 3. – P. 984-1026.

10. Global Health Economics. Shaping Health Policy in Low- and Middle-Income Countries / ed. Paul Revill [et al.] – New York, 2020. – 376 p. DOI.org/10.1142/11045.

11. Hamilton, W. N. The Fundamentals of Healthcare Administration: Navigating Challenges and Coordinating Care / W. N. Hamilton, D. J. Howell, Melissa Jordan. – New York : University of North Georgia Press, 2021. – 400 p.

12. Matthew, G. Transparency and negotiated prices: The value of information in hospital-supplier bargaining / G. Matthew, A. Swanson // Journal of Political Economy. – 2020. – Vol. 128, No. 4. – P. 1234-1268.

13. Neumann, P. J. Should a drug's value depend on the disease or population it treats? Insights from ICER's value assessments. Health Affairs blog [Electronic resource] / P. J. Neumann, M. C. Silver, J. T. Cohen. – Access mode: https://www.healthaffairs.org/do/10.1377/hblog20181105.38350/full/ (Date of access: 07.09.2023). – The name from the screen.

14. NUPh. Department of Organization and Economics of Pharmacy [Electronic resource] : official site. – Access mode: http://economica.nuph.edu.ua (Date of access: 02.10.2023). – The name from the screen.

15. Scientific library of NUPh [Electronic resource] : official site. – Access mode: https://lib.nuph.edu.ua (Date of access: 22.02.2023). – The name from the screen.

16. Sherman Folland. The economics of health and healthcare / Sherman Folland, Allen C. Goodman, Miron Stano. – 7th ed. – New York, 2020. – 621 p.

17. Song, P. H. Fundamentals of Healthcare Finance / P. H. Song, K. L. Reiter Gateway to Healthcare. – 4-th Edition Management. – New York, 2022. – 520 p.

18. World Health Organization [Electronic resource]. – Access mode: https://www.who.int/health-topics/health-workforce#tab=tab_1 (Date of access: 22.02.2023). – The name from the screen.

CONTENT

Конспект лекцій містить основні теми освітньої компоненти «Basics Economics in Healthcare», що пояснюють базові економічні категорії, закони і закономірності, принципи організації та побудови економічних відносин; охоплює питання фармацевтичних та медичних послуг, економічної оцінки в охороні здоров'я тощо. Конспект лекцій має зручну та логічну структуру, ілюстрації.

Для аудиторної та позааудиторної підготовки англомовних здобувачів вищої освіти спеціальності 226 «Фармація, промислова фармація» за першим (бакалаврським) рівнем вищої освіти.

Самостійне електронне навчальне видання

Немченко Алла Семенівна
Демченко Наталія Володимирівна

ОСНОВИ ЕКОНОМІКИ В ОХОРОНІ ЗДОРОВ'Я. КОНСПЕКТ ЛЕКЦІЙ

Навчальний посібник

Англійською мовою

Редакторка *О. Ю. Гурко*
Комп'ютерне верстання *О. М. Білинської*

Формат 60 × 90/16. Ум. друк. арк. 9,6.

Національний фармацевтичний університет
вул. Пушкінська, 53, м. Харків, 61002
Свідоцтво суб'єкта видавничої справи серії ДК № 3420 від 11.03.2009.