Value Judgements and Economic Evaluation

OBJECTIVES:

1. Describe the type of question that evaluative economics is intended to answer.
2. Describe what a value judgment is and how it can be used in evaluative economics.
3. Using assumed valuations by individuals for services and costs, identify an efficient level of output in any market.
4. Compare alternative delivery arrangements in terms of their efficiency.
5. Describe how a market for health insurance can be efficient when there is less than complete insurance coverage.
6. Describe the extra-welfarist approach to identifying optimal economic arrangements.
7. Define the concept of equity.
8. Identify several alternative measures of equity and explain how these can be applied to evaluate alternative modes of finance and care delivery.

12.1 INTRODUCTION

In this chapter we begin a different level of inquiry. In previous chapters we focused on the actual allocation of resources devoted to medical care. We were interested in explaining only the various allocations that might occur in different circumstances. We did not concern ourselves with whether any particular allocation was “good” or “acceptable” or “equitable,” to mention only a few of the terms we might use to label an allocation. In this chapter we begin the task of evaluating alternative possible allocations of resources. This task will lead us to such questions as whether totally free care can be judged “better” than the provision of medical care in a simple market. Or whether and in what sense a regulated system is preferable to an unregulated one. Many of these questions, it should be pointed out, are policy issues. Indeed, evaluative analysis forms the cornerstone of policy analysis, since the ultimate goal of policy is to bring about improvements in the use of resources.

Before undertaking evaluative analysis, we must lay the ground rules for conducting an evaluation. That is the mission of this chapter. In Section 12.2, the importance of having a recognizable and unvarying standard for gauging alterna-
tive allocations is discussed. The values that individual persons place on specific services can be used as the basis of a social evaluation. One procedure for building a social evaluation is discussed in Section 12.3. The standard that results from this procedure, which is used frequently by economists, is referred to as an efficiency criterion. Such a yardstick takes individuals' starting situations as given and therefore bypasses questions relating to equity and need as determined by clinical criteria. The application of efficiency criteria to evaluate the performance of the health insurance market is discussed in Section 12.4, and policy goals emanating from this efficiency analysis are presented in Section 12.5. The relevance of the efficiency criteria as the sole benchmark of resource allocation has been questioned by many observers. An alternative approach, called extra-welfarism, is presented in Section 12.6. Finally, alternative measures of equity are considered in Section 12.7.

12.2 VALUES AND STANDARDS IN ECONOMIC EVALUATION

Suppose we are faced with a situation in which A has a curable cancer but is receiving no medical care and B is healthy but is spending $4,000 on surgical services for a facial lift. Would this be an acceptable allocation of our medical resources? Many would say it is unfair, but scarcely anybody would take the trouble to set forth the basic standard being used to judge the situation. Suppose, instead, that it was necessary heart surgery B was receiving. Would this change one's evaluation of the situation? Would a different standard be used to gauge its fairness?

In our example, the resources are being allocated differently in the two situations. However, unless we had a standard that did not itself vary from situation to situation, we really could not compare the two situations. That is, without an independent scale of fairness or acceptability, we would not have a measure capable of assessing alternative allocations. This section presents a classification of available systems of standards, focusing on the bases on which standards may be formed.

For the purposes of economic evaluation, there are two ways of deriving a system of values and then developing a ranking of alternative uses of resources. In the first method, called delegatory or top-down, a value system is imposed on the members of society. For example, it might be imposed by a higher being, such as a deity; by an interpreter of the ultimate word, such as Moses or Mohammed; or by a dictator, who settles on some value system based on his or her values. Alternatively, someone can assume the mantle of spokesperson for society, proclaiming “society wants a decent standard of health for all” or some such alleged truth. Despite the nod toward democracy, any would-be ethical authority who chooses to speak for society without a mandate based on the views of individuals within the society is really imposing his or her own views on society.

The second method for deriving a system of values is called participatory or bottom-up. In this method, the views of all members of the community play a role. One assumption underlying this method is that everyone’s values must be taken into account in ranking alternative ways of using resources. Another assumption is that each individual is the best judge of his or her own welfare.

We now turn our focus to the value systems themselves. They vary tremendously, ranging from the very specific to the very vague. They can take the form of specific laws handed down by a deity or can be formulated in terms of general concepts such as fairness, liberty, and equality.
The field of health services analysis contains many examples of writers proposing value systems based on their own view of what seems desirable. For instance, some have posited a “right” to health or health care. One commentator used the principle of agape to derive this right (Outka 1974), whereas another appealed to a “strong sense in the population” that this right exists (Mechanic 1976).

Even assuming we could settle on a single value system, we would still face the problem of translating the chosen value system into a gauge or ranking scheme to assess alternative ways of using resources. This translation step can be controversial itself. Because any value system will be somewhat vague, different ranking schemes with very different implications can be derived from it. We would then run into the problem of which ranking scheme to choose. For example, the goal of “equality” can be interpreted in many ways—as equality of health status or equality of medical care utilization. We might decide it entails equal use of medical care for equal health status, with individuals who have poor health receiving more care than individuals who are basically healthy. This may seem plausible, but how do we decide how much more care people with poor health should get? Also, if the medical care given to those in poor health is not effective, should they still receive it?

The last step, after having decided on a ranking scheme, is to apply it to actual or proposed states of resource use (e.g., distributions of health care or levels of health) to determine their desirability from a policy standpoint.

It should be stressed that value systems imposed from above are not necessarily evil. The source of such a system may be a highly respected and beloved authority, and the system may contain laudatory ideals and translate into ranking schemes that seem reasonable and compassionate. Nevertheless, an imposed scheme is not built up from the values of the members of the society and therefore retains some degree of nonrepresentativeness.

In Section 12.3 and 12.4, a participatory system of evaluation is developed. This system, well known in economic circles as the Paretean system (named after the famous nineteenth-century sociologist Vilfredo Pareto), allows us to arrive at an optimum position through examining changes that could be made in resource allocations if we start from an initial position. This optimum holds only with reference to the initial starting point (i.e., the initial endowments each member of society possesses). We do not judge the starting point, which may or may not be fair, a consideration discussed in Section 12.5.

### 12.3 EFFICIENT OUTPUT LEVELS

#### 12.3.1 Individual Valuations of Commodities or Activities

If we accept individuals’ own valuations as the best indicators of their own welfare, we must then determine, at least in principle, what these valuations might be. Since our analysis is concerned with specific commodities, our task is simplified somewhat. We need only determine individuals’ valuations with respect to those commodities with which we are concerned.

Economists have developed a hypothesis regarding an individual’s valuation of units of a specific commodity. The hypothesis, which is based on our demand analysis, is that the more of any commodity the individual has, the less successive units of the commodity will be worth to him or her (as compared with other commodities). The analysis can be recast using money as the basic unit of value. To
do this, we must assume that money is itself of constant value. That is, if an individual gives up $2, that $2 will always represent the same loss to the individual however much income he or she has. This assumption will hold, at least partially, if the outlay for the commodity in question is a reasonably small portion of the individual's total budget.

If an individual has an income of $10,000, spending $100 or $150 on a commodity is unlikely to cause the valuation of each dollar to change for the individual. However, as the amount that must be given up to obtain a commodity becomes very large relative to income, the utility of or the subjective valuation placed on the marginal dollar will change. We are making the assumption in this section that it does not. We should note that this is a different assumption from that made earlier in the discussion of health insurance demand (Chapter 10).

Given the assumption that money income has a constant value for individuals for all relevant ranges of expenditures, we can specify individual valuations of successive units of a commodity in terms of money. These valuations, it must be stressed, are the individuals' own evaluations of specific units of the commodity, and they qualify on participatory grounds for inclusion into our overall participatory social evaluation.

12.3.2 Values in a Selfish Market

To simplify our analysis, let us assume initially that there are two individuals in our market, A and B. Each has a specific schedule of valuations for his or her own consumption of medical care. Let us refer to these valuations as marginal valuations (MVs). A marginal valuation is defined as the extra amount of money an individual would be willing to pay for an additional unit of a commodity. Thus, an MV is a measure of what an extra unit of the commodity is worth to the individual in money terms.

In our initial analysis, both A and B derive satisfaction or value from their own consumption of medical services, and theirs is the only satisfaction that anyone in society gets from their consumption. A places a marginal value of $80 on his first unit consumed, $70 on his second, and so on, as seen in Columns 1 and 2 in Table 12-1. Note that the marginal values placed by each individual on successive units of medical care consumed diminish. Recall from Chapter 3 that all other factors, such as health status, income, and wealth, are held constant (i.e., the initial values of these variables are held constant). For purposes of social evaluation, then, we have a measure of the social worth of A's consumption of medical care (since no one else values this care other than A himself).

The assumed relation between marginal value and quantity consumed can be presented geometrically. In Figure 12-1, the curve $MV_a$ represents A's marginal valuation of successive units of medical care. It is assumed, for ease of geometric exposition, that the units of medical care can be made very small so that the $MV$ curve becomes smooth. A's valuation of his own consumption is referred to as the private (or internal) valuation of his consumption. On the assumption that no one else cares about A's consumption, his private valuation is the same as the social valuation (the total value placed on A's consumption by all of society).

Similarly, we present the private valuations of B in Table 12-1 and geometrically as $MV_b$ in Figure 12-1. For whatever reason (she is poorer, more healthy, or less well educated), B places a lower value on each unit of health care than does A.
Table 12–1 Values and Costs of Medical Care

<table>
<thead>
<tr>
<th>Quantity Consumed by A ($Q_a$)</th>
<th>Quantity Marginal Valuation by A ($M_{V_a}$)</th>
<th>Quantity Consumed by B ($Q_b$)</th>
<th>Quantity Marginal Valuation by B ($M_{V_b}$)</th>
<th>Quantity Consumed by A and B</th>
<th>Social Marginal Value of Consumption</th>
<th>Marginal Cost of Output at Consumption Level $Q_a + Q_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$80$</td>
<td>0</td>
<td>$0$</td>
<td>1</td>
<td>$80$</td>
<td>$35$</td>
</tr>
<tr>
<td>2</td>
<td>$70$</td>
<td>0</td>
<td>$0$</td>
<td>2</td>
<td>$70$</td>
<td>$35$</td>
</tr>
<tr>
<td>3</td>
<td>$60$</td>
<td>0</td>
<td>$0$</td>
<td>3</td>
<td>$60$</td>
<td>$35$</td>
</tr>
<tr>
<td>4</td>
<td>$50$</td>
<td>1</td>
<td>$50$</td>
<td>5</td>
<td>$50$</td>
<td>$35$</td>
</tr>
<tr>
<td>5</td>
<td>$40$</td>
<td>2</td>
<td>$40$</td>
<td>7</td>
<td>$40$</td>
<td>$35$</td>
</tr>
<tr>
<td>6</td>
<td>$30$</td>
<td>3</td>
<td>$30$</td>
<td>9</td>
<td>$30$</td>
<td>$35$</td>
</tr>
<tr>
<td>7</td>
<td>$20$</td>
<td>4</td>
<td>$20$</td>
<td>11</td>
<td>$20$</td>
<td>$35$</td>
</tr>
<tr>
<td>8</td>
<td>$10$</td>
<td>5</td>
<td>$10$</td>
<td>13</td>
<td>$10$</td>
<td>$35$</td>
</tr>
<tr>
<td>9</td>
<td>$0$</td>
<td>6</td>
<td>$0$</td>
<td>15</td>
<td>$0$</td>
<td>$35$</td>
</tr>
</tbody>
</table>

Indeed, her first unit has an $M_V$ of $50$, her second has an $M_V$ of $40$, and so on. These valuations might seem low to us, but since B is the ultimate judge of her own welfare, we cannot question these valuations: they are simply part of the data.

According to our assumption, A and B are the only members of society who participate in the medical care market. The marginal social valuations of medical care coincide with the marginal private valuations. Column 5 of Table 12–1 lists the

![Graph](image)

**Figure 12–1** Representation of efficient output level. Individuals A and B have private marginal valuations for medical care ($M_{V_a}$ and $M_{V_b}$ respectively). Using these, we calculate the marginal social value ($MSV$) curve, which relates aggregate quantity to each individual's valuation. $MC$ is the marginal social cost of medical care. The efficient level of output is that quantity at which $MSV$ equals the $MC$.  


aggregated quantities that correspond to each level of MV. For example, at an aggregate quantity of five units of medical care (four used by A and one by B), each consumer's marginal value will be $50. If seven units were consumed (five by A and two by B), each individual's MV would be $40. We now have hypothesized how much each additional unit of medical care is worth to each participant. Furthermore, we have derived an aggregate-level relationship between the quantity of medical care and the marginal value to each individual if he or she was consuming at the level of consumption indicated by the MV curve (Figure 12–1). This aggregate curve, called the MSV (marginal social value) curve, shows the value to each member of the market if all individuals are consuming at the levels indicated by the curve. Because B does not have an MV above $60, for values above $60 the MSV curve coincides with A's MV curve.

An implicit assumption of our analysis is that consumer valuations are expressed in terms of a commodity, medical care. But medical care may not be valued for its own sake (except, perhaps, by a hypochondriac); it is usually health that is valued. In fact, each consumer's MV is made up of two components: an MV for health (termed H) and the marginal productivity of an additional unit of medical care (M) in producing health (ΔH/ΔM). Thus, the valuation of medical care is derivative, stemming from the two components.

We come next to the cost of producing medical care. Our initial assumption here is that each level of output is being produced at the minimum cost. This assumption is sometimes referred to as the technical efficiency assumption. It implies that, given production conditions and input prices, the lowest cost combination of inputs is used at any output level. In Column 7 of Table 12–1 and in Figure 12–1, we show the minimum marginal cost at which providers can produce medical care. We assume that this minimum marginal cost remains constant at $35 per unit as output increases. Note that the MC is the additional cost per unit of care; each extra unit costs $35 to produce.

One interpretation of MC is that it is the amount of money that must be paid to the inputs to hire them away from the next highest valued use. If medical care was not produced, something else of value to consumers would be. We can assume, then, that the MC is the amount we would have to pay the resources to induce them not to produce that something else. This approach allows us to put a value on unpaid resources that otherwise would appear to be "valueless" or "free." Thus the MC is marginally above (and approximates) the value that someone else would have placed on these resources in an alternative use. Viewing the MC in this way means that it is essentially the opportunity cost of the resources used (the value that users of other commodities would have placed on them).

12.3.3 The Socially Optimum Quantity of Medical Care

The next step in our analysis involves the definition and identification of desirable or optimum resource allocations. Since our method of evaluation is participatory, we need to identify allocations of resources that would be considered better than alternatives by all members of the community. As will be seen presently, it is possible using a participatory method to rank some allocations as superior to others, although we cannot compare every conceivable situation. Our criterion is this: the resources must be used in a way that maximizes social value. That is, if the
resources are distributed in such a way that consumers are willing to pay the most for them, then output will be at the "right" or economically efficient level.

Using the valuations of A and B and the MC of medical care, we will be at a socially optimal (or economically efficient) level of output if the MVs of A and B equal the MC (i.e., \( MV_a = MV_b = MC \)). If output is at a level where the MVs are greater than MC, say, at an aggregate quantity of 3 in Table 12–1, then an expansion of output to 5 (an increase of 1 for A and B each) would have an MC per unit of output of $35 but would yield $50 extra in value to A and B each. Similarly, if the MC is greater than the MSV, this indicates that resources are worth more elsewhere, and so output should fall. In Figure 12–1, the optimal level of medical care is 7 units. Given our assumptions, this is how much medical care should be produced. This measure of efficiency—the distribution of output based on utility—is called allocative efficiency.

In reality, in a medical care market too much or too little as well as just enough medical care could be produced. Too much could be produced if the government had a policy of financing medical care and giving it away for free. At a zero price, demand will be at 15 units (where the MVs are zero); the MC of additional units will be well above this if the government is willing to ensure that all that is demanded is provided. The financing of the program could be through taxes. However, by meeting all demands, the government is clearly providing too much.

On the other hand, the market may provide too little. If medical care was in the hands of a monopolist, the monopolist would set a price well above that where \( MV = MC \). If the price was $55, then three units in total would be demanded (all by A). Here the market would be producing too little care.

In addition, the optimum level of resource use could result in little or even no use of medical care by some individuals. The height of the MV curve, which is in effect a demand curve, will depend on health status, wealth, income, and so on. Poor people (e.g., B) may have low MVs. Indeed, if the MC was higher than in our example, a socially optimal quantity of output would be perfectly consistent with no consumption of medical care by B. (This is true, even though B may have poor health.) One might argue that this is unfair, and, indeed, depending on one’s definition of fairness, it might well be. It should be recognized, however, that the root cause of the inequitable distribution of medical care is the inequitable distribution of wealth. A higher income for B would mean higher demand and MV curves for medical care. Of course, as far as the notion of economic efficiency is concerned, initial wealth and income levels for each individual are given. A redistribution of income or wealth among individuals might seem fair to many observers, but it would not be evaluated within the bounds of the present notion of economic efficiency.

12.3.4 Optimal Output with Altruism

To preserve the present notion of economic efficiency and to extend it to cover some distributinal issues, an analysis has been developed to allow for the concern of some individuals for the low medical care consumption levels of others. Let us extend the previous example to allow for A’s external demand for B’s consumption of medical care. From A’s viewpoint, it may well be that B has a level of consumption of medical care that is too low. If this is the case, we must find some representation
of the value to A of B's medical care consumption. It is likely, of course, that A's concern for B's medical care consumption is not unlimited. A is concerned, but only up to a point, for A has other private and public concerns as well. In fact, as seen in Section 4.3, A's valuation of B's medical care consumption can be treated as any other commodity; the more B consumes, the less the marginal value to A of an additional unit. In Table 12-2, A's MV for B's consumption is $30 for the first unit, $20 for the second, and so on. In Figure 12-2, this external MV curve is shown as \( MV_a^b \).

It may seem strange that A's altruistic concern for B's welfare can be translated into mercenary terms and be given a money measure. Our ability to do this rests on the assumption that commodities are scarce and A must make some choices at the margin. Even if A decided to give all his money away and use none of it for his family or own personal use, there would still be hard decisions to make. Should the money be donated to the cancer society or heart association? Should the money go toward the preservation of Newfoundland seals or bald eagles? Depending on their tastes, even the most altruistic of people must make choices regarding scarcity, and our analysis is merely a formalization of this fact. Of course, most people will engage in private consumption as well as altruistic consumption; their values can be presented by marginal valuation curves for both types of activities. The benefits to be obtained from others' consumption will be termed external benefits, and the values that people place on these benefits will be termed external values.

We can now arrive at a measure of the value that society places on B's medical care. This value can be called a social value, and it is made up of all individuals' private and external values for the specific commodity. Thus the marginal social value for B's consumption of medical care can be obtained by adding up the values both individuals place on each successive unit of medical care that B might consume. In Table 12-2, society has a marginal valuation of $80 for the first unit of B's medical care (equal to the sum of \( MV_b^a \) and \( MV_a^b \)), $60 for the second, and so on. These valuations are shown in Figure 12-2 as MSV, which is the vertical sum of \( MV_b^a \) and \( MV_a^b \). By vertical sum we mean that each unit of B's consumption has a value to society (A and B) greater than the value placed on it by B alone. Because of this "public" dimension, we sum all values placed on each unit of B's consumption. Since each member's valuation of the commodity is measured along the vertical, or cost, axis, the summation of all members' valuations of this commodity is therefore a vertical sum.

The marginal valuation curve facing the market for medical care for A and B is MSV, which shows the quantity for all individuals at alternative MSVs for each

<table>
<thead>
<tr>
<th>Quantity Consumed by B</th>
<th>Marginal Value to B of Own Consumption (( MV_b^a ))</th>
<th>Marginal Value to A of B's Consumption (( MV_a^b ))</th>
<th>Marginal Social Value of B's Consumption (( MV_b^a + MV_a^b ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>30</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 12–2 Representation of efficient output level. Individuals A and B have private marginal valuations for medical care (MV_A and MV_{b'}, respectively); in addition, individual A places an external value on B's consumption. The marginal social value of B's consumption is the sum of the values placed on B's consumption by both A and B. The MSV of all medical care also reflects this externality. The MC is the marginal social cost of medical care. The efficient level of output is where MSV = MC (at approximately eight units).

individual. This curve is much like the MSV curve in Figure 12–1, except it incorporates A's valuation of B's consumption along with the private MVs.

The socially optimum level of output is similarly interpreted; the output is optimal at the quantity where the MSV for all individuals equals the MC. In Figure 12–2, the optimum level of output is eight units of medical care. This optimum quantity incorporates each individual's private valuations as well as any external valuations for the poor, the needy, the sick, and so on. The optimum quantity that incorporates the external concerns of A is greater than the optimum quantity if only selfish concerns exist (see Figure 12–1). However, these outcomes are results of the data, and it may well be that B's optimal consumption is still at a low level.

The results of our extended analysis are consistent with some kind of transfer of funds from A to B for the purposes of increasing B's consumption of medical care. However, the analysis does not say what kind of transfer should take place. It may be voluntary (e.g., charitable donations given by A directly to B or to some providing agency) or tax based (e.g., taxes levied on A might be used to reimburse providers). Although if taxation is used to raise funds, our analysis implies that it is voluntarily accepted by A. In either case, the optimal solution allows for some transfer, but it should be stressed that a transfer can be too much or too little. The government can
over- or underprovide, based on A's criteria. All that our analysis shows is that some transfer is consistent with economic efficiency.

12.3.5 Alternative Delivery Arrangements

Now that we have identified an ideal or efficient output level, we can look at alternative delivery arrangements to see how they compare to the ideal. That is, we can determine whether expected output under the alternatives is too little, just enough, or too much.

12.3.5.1 "Free" and Unlimited Care

First, assume that B is given all the medical care for free that she can consume. As column three of Table 12-1 indicates, she would choose to consume six units of output. The social optimum is eight units for A and B, and the MC at this quantity is $35. Optimally, B should consume three (i.e., where $MSV_b = MC$). For every unit B consumes beyond three, the value of B's consumption is less than the cost to society (everyone). Since someone must bear the burden of this care, and since $MC$ exceeds $MSV$ for all units beyond three, there is a net social loss for these units. B gains handsomely (i.e., her private benefits exceed her private costs), but overall this type of arrangement may lead to a great deal of medical care being consumed with very little value attached to it.

12.3.5.2 Competitive Market, No Philanthropy

Let us look at another arrangement, that of a competitive market with no philanthropy or government programs. Recall from Chapter 8 that equilibrium in a competitive market will occur where marginal private cost equals price. In the example here, A will consume the right amount for himself, but B will not. B's consumption will be less than the socially efficient amount because all society would have been willing to pay more for the first four units of B's consumption than the marginal cost. The competitive market does not provide a mechanism to express A's external demand for B's care. A freely operating competitive market with no philanthropy will yield less than the optimal level of output when externalities would have justified a larger output. As for a monopolistic market, the output of such a market will be less than the output of a competitive market, which means it will be even further below the optimal amount.

12.3.5.3 Competitive Market with Philanthropy

It has been contended that a competitive market even with philanthropy will not produce the optimal amount of output. To understand why, consider a situation in which there are many donors of medical care, each of whom places a value on the consumption of medical care by the needy. In this case, some social arrangement must be found for ensuring that the values of these donors will be expressed in the market. If each of these potential donors offers to give what the output is worth to him or her, the social value will equal the sum of the private values. However, if each donor feels that the others will also give, he or she might give less, hoping to get a "free ride," that is, gain the benefit of the others' donations
while giving less. It is in the interests of each private donor to initially offer less than the value he or she places on the output in the hope that someone else will pay the tab. If everyone behaves in this way, the total amount given in philanthropy will be less than the socially optimal amount. Analysts who accept the efficiency criterion frequently justify compulsory government programs on the basis that they make everyone pay what the programs are worth. Of course, it is difficult to decide how much a program would be worth to each taxpayer, since the individual still has an incentive to understate the value of the program to him- or herself.

Even accepting this justification for government programs, we still must discover whether there exists an arrangement that will lead to the correct amount of medical care being utilized. As can be seen in Figure 12–2, if B was offered subsidized medical care, the efficient amount of medical care would be utilized. In this case, a charge of $30 per unit of medical care to B would lead to B’s consumption of the optimal quantity—three units. The rest of society must now pick up the remainder of the tab. Since the total cost to all members of society of medical care consumed by B is $105 and since B will pay $90 of this, some arrangements must be made to collect the remaining $15 from the rest of society. This can be done in the form of taxes. Various arrangements are discussed in the next paragraphs.

We can conclude from our analysis that some form of cost-sharing arrangement can lead to the provision of an optimal or efficient amount of the product. However, other arrangements can also be efficient. One is to have needy individuals pay nothing and to impose some form of rationing. In practice, this type of arrangement requires that the rationing system used must produce the efficient outcome, and such systems are difficult to design and operate. Our analysis can also be extended to a case in which the needy individuals have different levels of income. If their demands differ because of these income levels, a system of variable subsidies tailored to income levels could be designed to have each member consume the right level of output (Pauly 1972).

What is critical in translating the preceding analysis into a policy prescription is a clear conception of what the external demands might be in actuality. Assuming that external demands for the medical care of some groups do exist and are significant, it is essential that we pin down exactly what services these external demands are for. If they are for good health, for example, then the external demanders (the A’s in our analysis) may demand preventive care for consumption by the potential recipients of aid (the B’s). The demands may be much more specific than that, however. The demanders might show concern only for individuals who have catastrophic illnesses requiring large financial outlays. In this case, they will not want to pay for the medical care of needy individuals who have sore throats, ingrown toenails, or acne. We know very little about the nature of medical care externalities (external demands). From an efficiency point of view, however, it is necessary to know what the external demanders are concerned about before we design a delivery system that will incorporate these externalities.

Assuming that we have identified the nature of the external demands, we can then use the preceding analysis to answer our questions, as long as we have the goal of efficiency in mind. Once the demands have been pinpointed, the types of health care that might improve the situation and the potential recipients can be identified. The consumer’s portion of cost sharing should be designed to ensure that there is no overuse, which is defined as any quantity beyond which marginal social benefits are
less than marginal social costs. The reimbursement mechanism chosen should lead to the least cost output.

12.4 OPTIMAL HEALTH INSURANCE

The provision of health insurance requires resources and incurs costs. In the same way that there is an optimal quantity of medical care, there is an optimal degree of insurance coverage (see Section 10.2.1). We will assume that all individuals are the same in all respects except one—the amount they must pay to obtain insurance.

Let us assume that there are 900 individuals (the number is not important) who are members of a large group and 100 individuals who are members of a small group. All individuals have an initial level of wealth of $1,000. There is a likelihood of 10 percent that each individual will get sick (i.e., 10 percent of the group will get sick). For those individuals who do get sick, the medical costs are $200 per patient. The utility function for each member (all have the same tastes) is as shown in Table 10–1. This utility function can be interpreted as a measure of "consumer welfare." With regard to the supply side of the market, we assume that there is one insurer who provides insurance at cost. The loading cost to the insurer of a large group policy is $30 whereas the cost for a small group policy is $60. Our objective is to maximize the overall utility of all members without detracting from that of any single member. This is the Paretean criterion.

The framework we will use focuses on consumer welfare (utility). In general, we can assume that consumer welfare is maximized by shifting the risk onto the insurer whenever the expected utility with insurance is greater than the utility in the absence of insurance. The postinsurance utility is the net of the economic cost of accepting the risk. Therefore, utility (welfare) is maximized whenever the risk is appropriately shifted.

In our analysis, there are two groups of individuals. Each individual faces an expected loss of $20 (i.e., 10 percent of $200), and each can obtain insurance at a cost that includes the expected loss ($20) plus the appropriate loading cost. For members of the large group, the full premium, including the loading charge, is $50. For members of the smaller group, the premium is $80. For members of the large group, there is a utility or welfare gain by shifting the risk: at a cost of $50, the utility will be 97.0 units, which exceeds the expected utility of not insuring, which is 95.8 units. There is a social gain from shifting the risk. The same is not true for the members of the smaller group. Since the cost of insurance for them is $80, they would be better off to remain uninsured. This would be true even if the cost of insurance for the smaller group was subsidized (i.e., if someone else paid part or all of the premiums). This is because we are using the criterion of social efficiency rather than individual efficiency. When we recognize that there is a social cost of insuring, then we must also recognize that there is an optimal degree of insurance coverage. This optimal degree may be zero if the arrangements for providing insurance are too costly.

We must also acknowledge that consumers may vary in many respects, including the following: risk of illness, income or wealth level, degree of risk aversion, and circumstances affecting the cost of illness. As each varies, the utility gain from shifting the risk of incurring medical expenses will also change. For example, individuals with a high risk of illness will gain more in utility terms from shifting
their risk than individuals with a low risk of illness. Thus, a situation in which individuals who are less healthy have greater insurance coverage could be an optimal situation. That is, variations in insurance coverage between individuals can be economically efficient.

There is a confounding factor in this analysis—moral hazard. There can be a net welfare gain resulting from the shifting of risk. Once the risk is shifted, the out-of-pocket price of medical care to the consumer falls. If there is any elasticity of demand for medical care, then moral hazard will come into play and the quantity demanded of health care will increase. If the out-of-pocket price of medical care is low enough, the individual might consume care up to the point where \( MC > MV \). There is a net welfare loss in the medical care market that occurs when the individual is ill. There are, then, two welfare effects of insurance: the welfare gain from shifting the risk and the welfare loss from consuming beyond the optimal point when the individual is ill. True optimality requires that we consider both effects together (Gianfrancesco 1978). Usually investigators focus on the insurance market (Gianfrancesco 1983; Pauly 1990) or the medical care market (Pauly 1972) in isolation from one another.

12.5 EXTRA-WELFARISM

The framework we have used until now includes a number of value judgments and principles. A key principle is that each person is the best judge of his or her own welfare. Welfare, in this framework, depends exclusively on the utility of goods and services as valued by the individuals. If there is any “public” component of goods and services, it is introduced through external demand, which is the value some people place on other people’s consumption. Beyond this, there is no justification for publicly provided health care that can be derived from the Paretean welfare framework.

The Paretean framework has come under criticism in recent years on the grounds that it does not include all that people value in life (Culyer 1990; Rice 1992). There are other sources of personal well-being besides goods and services. Many of these other sources of well-being are embodied in the characteristics of people rather than the characteristics of the goods and services that people consume. People value mobility, absence from pain, and absence from distress—and they value these for other people as well as for themselves. While it is true that there are commodities (including medical care) that are linked to these more ultimate sources of well-being, there is no automatic link between them. Consequently, a social evaluation based on commodities consumed and nothing else appears much too narrow.

“Health” is often viewed as a composite of characteristics of people, such as mobility, absence of distress, and so forth. A number of economists have asserted that health is important not only because we want it for ourselves. They regard health as one of several entities that “society” recognizes should be made available to everyone (Culyer 1993) regardless of willingness to pay. This position has often appeared in the health care literature (Fein 1972; Outka 1971). If health really is a socially recognized good, then health services cannot be evaluated strictly in terms of their market value. In particular, the distribution of health services must be evaluated on a social basis.
The researchers who hold this position largely avoid the question of who the judge of welfare will be, a question directly addressed in the Paretean framework. They merely assert that some decision maker, chosen (or elected) by society, should be responsible for conducting the evaluation. Thus we are no longer clear, in this extra-welfarist viewpoint, who the judge of welfare is. Indeed, extra-welfarism is consistent with the use of any social judge other than the consumers; the approach merely posits that there are some entities whose social value is determined outside of the consumers themselves. The role of economists is to act as advisors for the distributive organization and uncover the implications of incorporating efficiency and other objectives into the economic analysis. It should be pointed out that people’s direct evaluations of their health services can be included in the extra-welfarist economic calculus, as can other (nondirect) evaluations of their health care.

The extra-welfarist position is concerned with how health is distributed among all members of society. Whoever the judge of well-being becomes (the government, a community league, etc.), value judgments must still be made in order to decide how to distribute health services and health. One way to operationalize the extra-welfarist approach (i.e., turn it into an evaluative tool) is to provisionally accept the principle that health care should be distributed according to “need.” If need is defined as the ability to benefit from health services (Culyer 1995), then the “decision maker” is faced with the question of how to allocate health services so as to enhance or preserve different individuals’ health status. Even if this approach evades the issue (or at least leaves the issue open) of who is to decide on the distribution, it helps make explicit the wide array of distributions that are possible (using the principle of need and other principles as well).

Figure 12-3 is a graph that shows the health of two individuals, A and B, measured along the two axes (Wagstaff 1991). Let us make the following (non-value-laden) assumptions: individual A has a self-assessed health status of \( h_a \) and individual B has a self-assessed health status of \( h_b \). The health status of both can be improved, but there is a limit. Curve \( H \) shows the maximum amount of health that can be produced with the resources available for health care (assumed to be fixed for society as a whole). More health can be produced for A, but only at the expense of resources and health for B. With available resources, A’s health can be increased up to \( h_a \) (with no change in B’s health) and B’s health can be increased up to \( h_b \) (with no change in A’s health). The exact shape of the \( H \) curve will depend on how effective the additional resources are in improving each individual’s health. If very little extra can be done to improve B’s health, then the curve will be steeply sloped. Our curve shows that more can be done for both.

Mentioned above was the principle that health resources should be distributed according to need. There are a number of different ways to express “need.”

- **Equal health status.** One value judgment is to allocate resources so that everyone ends up with an equal level of health. If this principle is used, then more resources must be provided to A to ensure that in the end both A and B are equally healthy. Equal health implies that each person is on a 45° line from the origin.
- **Maximizing total health regardless of its distribution.** In order to implement this criterion, we must know the trade-off in health status between the two persons. If the health transformation curve, \( H \), favors person A, then resources will be more productive in improving A’s health status rather than B’s. An optimal point will
Figure 12-3 Potential health status of two individuals. The health status of A and the health status of B are currently at $h_1$ and $h_2$. Through the expending of more resources, their health status can move up to $h_x$ and $h_y$ respectively. However, since the available resources are finite, the limit of improvement for both individuals combined is shown by curve $H$.

The usefulness of the extra-welfarist approach is that it allows us to go further in exploring resource allocation than the Paretean or welfarist position, and if society places special importance on characteristics such as health, then alternative distributions of health care resources need extremely careful evaluation.

### 12.6 CONCEPTS OF EQUITY

Distributional equity is important in analyzing both access to and consumption of medical care (e.g., differences in utilization among groups) and its financing (e.g., differences in payments), and so it is essential to have measures of equity. We focus here on three types of distributional equity: intergenerational equity, vertical equity, and horizontal equity (Long and Smeeding 1984).
Intergenerational equity, in a financial context, concerns the distribution of payments among different generations. For example, if we divide up the population into retirees (who are generally over 65 and eligible for Medicare benefits), those of working age (say, those 18 to 64), and others, our classification scheme could be regarded as dividing the population along generational lines. Since the Medicare hospital insurance program is financed largely through the flow of payroll taxes into the Hospital Insurance Trust Fund, these taxes will be borne largely by individuals in the working-age group. In other words, the working-age generation is largely financing the care of the generation of retirees. The equity implications of this kind of tax are very different those of the tax used to expand the Medicare program’s benefits in 1988. This latter was a 15-percent tax on the taxable income of the retirees, and it proved to be so unpopular that the program expansion was repealed by Congress. This type of tax involves a minimal intergenerational transfer of funds.

Current employment-based private health insurance provides another example of intergenerational transfer. All employees pay a similar health insurance premium, which is based on the average utilization pattern for all workers. If workers were rated separately by age group, according to insurance principles younger workers would have a lower premium than older workers because their utilization is less. The financing method of charging everyone in the plan the same rate (community rating) is in effect intergenerationally inequitable.

The second type of equity, vertical equity, concerns the economic burden experienced by different income groups. For example, imagine we have three income groups: those who make under $20,000, those who make from $20,000 to $40,000, and those who make over $40,000. A tax is progressive if members of a higher income group pay a larger portion of their income in tax than those with a lower income, it is neutral if the portion is the same for all groups, and it is regressive if members of a lower income group pay a higher portion of their income in tax than those with a higher income.

An example of vertical inequity would be a flat tax charged to all individuals regardless of their income level. A fixed premium for Medicare enrollees is such a tax. Lower income groups pay the same rate as higher income groups do, and this premium amounts to a higher portion of their income.

Horizontal equity concerns the degree to which equals are taxed equally. An example of a horizontally inequitable tax is a tax on specific commodities such as alcohol, tobacco products, and hospital care. Consumption or sales taxes on the former two products fall on groups who use these products more heavily. Such taxes have been a popular means of financing health insurance programs for indigents. Even though these taxes are horizontally inequitable, it has been argued that, since these individuals are likely to be less healthy and use the health care system more, they should pay higher taxes. That is, if not just taxes but rather the net of health care services minus taxes paid is considered, then horizontal inequity is not present. Another type of tax that has been recommended as a way to pay for medical care for indigents is a tax on hospital admissions. Such a tax will also be inequitable, though to a large extent it will be less visible, because it will be passed on to the third parties who reimburse the providers. (Of course, the insurers, in turn, will pass the tax on by charging higher health insurance premiums.)
12.7 GOALS OF HEALTH POLICY

There are a number of different ways that the goals of health policy can be articulated. At one level, one can articulate a set of environmental conditions that will allow for a smoother operating policy. For example, many feel that if consumers are given a range of health plans to select from and the freedom to choose among those plans, then social goals will be forwarded. At a lower, operational level, there are goals that deal with operating performance, such as efficiency, equity, and public financial constraints. Finally, social goals can be articulated in terms of the health outcomes of a group of individuals. In order for the policy goal to be operational, policies will still have to specify how each person’s health status is to be included (e.g., whether everyone is counted the same). In addition, the policy makers should take into account the economic aspects of the policy goals. There are insufficient resources to allow everyone to maximize their health status; therefore, goals have to be set that allow one to rank social measures of health status resulting from different policies.

One might combine policy goals from the different categories. For example, a policy maker might deem equity of resource use to be important. However, once individuals have equal resources available to them (e.g., though spending vouchers), the policy maker might value freedom of choice to allow individuals to select those types of care that they feel would best suit them. As another example, a policy maker might value a population’s health status highly but also want to ensure technical efficiency is achieved. The policy maker would select policies that would forward both objectives. Below we discuss briefly each of the goals.

12.7.1 Environmental Conditions

A market is an institution in which free choice is exercised by the participants, who are in pursuit of their own well-being. A social goal that falls under this category is allowing persons the freedom to seek care from whatever health plan they want to join. One might also suggest that anyone who wants to provide health services may do so. Generally, choice in health care markets is not extended to anyone who wants to be a provider. This is a reflection of the fact that other policy goals, such as the protection of quality, are being sought.

12.7.2 Efficiency

In order to achieve economic efficiency, (1) demand must be at an appropriate level, neither too restrictive (e.g., because of monopolistic prices) nor too low (such that there be an excessive demand), and (2) providers must produce an adequate supply of services (3) at an appropriate quality level and (4) at a low cost of production.

12.7.2.1 Demand Barriers

Demand barriers are impediments obstructing the reception of care. Within the context of our present model, price is the prime impediment. One can encourage additional care demanded by lowering the direct price through the purchase of insurance, public programs, or charity. To the degree that additional medical care
utilization is thought to be desirable, the extent of financial demand barriers can be measured by the availability of insurance or the direct price faced by individuals.

However, money price is not the only factor related to demand barriers. Waiting costs and travel costs can also obstruct access of care. If medical care consumption is to be encouraged, these costs must be addressed, either through subsidies, relocating facilities to lower travel time and expenses, or expanding facilities and increasing operating hours to decrease waiting time. However, it should be remembered that, from an efficiency standpoint, demand can be too great as well as insufficient.

12.7.2.2 Adequacy of Supply

Adequacy of supply refers to the availability of sufficient resources to provide care at the efficient level (given the level of quality). Adequacy of supply depends on the incentive (reimbursement) system developed, the level of reimbursement, and the adequacy of funds.

12.7.2.3 Technical Efficiency

Technical efficiency is a measure of the cost of producing a given level and quality of output. Technical efficiency is usually expressed in terms of money costs, but care must be taken when comparing costs between facilities to be sure that all other factors (e.g., quality, input prices, and case mix) have been accounted for.

12.7.2.4 Quality of Care

On the assumption that quality is not free, it costs more to achieve a higher quality of care. Therefore quality, like any other characteristic of output, can come in too great or too little a quantity. Quality of care is an often cited policy goal in health care. Regulation and licensing of professionals are often enacted in the name of the protection of quality of care.

12.7.3 Equity

Equity is a very broad concept. We consider two aspects, equity of utilization and equity of finance.

12.7.3.1 Equity of Utilization

A service can be provided efficiently, and yet some individuals who could benefit from more of it simply cannot pay for it. Society, or policy makers, can set utilization goals above those that are provided in a market situation. In this case, the direct price or other barriers to care must be removed. Thus one may lose some efficiency in order to attain a higher degree of equity.

12.7.3.2 Equity of Finance

Equity of finance can refer to direct out-of-pocket prices as well as taxation and premiums, factors that may not directly affect utilization. Health care premiums
may be deemed too low, in which case some individuals would be viewed as not paying a fair portion of the cost of medical care. Equity of finance would call for an increase in these premiums.

12.7.4 Public Financial Constraints

Strictly speaking, the government budget does not fall within the scope of our model. Of course, a transfer of funds from A to B is consistent with a tax on A by a government body and subsequent expenditures on medical care for B. But the model says nothing about the size of the tax, the expenditure, or the difference (the contribution to the deficit). In recent years, however, the budget deficit and public spending have come under a great deal of scrutiny, and cutbacks in government programs have been widespread. Typically the rationale for cutting a program’s expenditures is not lack of worthiness of the program but the program’s contribution to the overall budget deficit. To the extent that cutbacks can be achieved merely through increases in technical efficiency, true savings are provided to society and there are gains in social efficiency. However, cutbacks may also result in reduced supply. This is not necessarily bad if output was greater than the socially optimum level to begin with. However, if the initial output was at the socially optimum level or below it, cutbacks will lead to reductions in social efficiency because the value of the output that is lost is greater than the savings resulting from the cutbacks.

12.7.5 Health Status of the Population

In recent years, many investigators have focused on population-based measures of health status as a goal of policy. Many of the other goals can be viewed as leading to better health, and so these investigators focus on direct measures of health as a policy objective. Of course, the costs of achieving various levels of health status must be considered as well. Thus, one is faced with the constrained objective of maximizing population health subject to resource constraints. Additionally, focusing on this goal does not do away with equity considerations. Once we have more than one individual whose health is measured, we face the problem of how to add up the health status of all the individuals. These problems have been discussed in this chapter, and they need to be addressed in any policy consideration.