Cost-effectiveness analysis (CEA) and cost-utility analysis (CUA) are commonly used alternatives to CBA. These methods are potentially useful when analysts seek efficient policies but face certain constraints that prevent them from doing CBA. Three constraints are common: First, analysts may be unwilling or unable to monetize the most important policy impact. This constraint arises most frequently in the evaluation of alternative policies that save lives: Many people are willing to predict the numbers of lives saved by alternative programs but are unwilling to place a dollar value on a life saved. The second constraint arises when analysts recognize that a particular effectiveness measure does not capture all of the social benefits of each alternative, and some of these other social benefits are difficult to monetize. In using CBA, analysts face the burden of monetizing all impacts. If the effectiveness measure captures "most" of the benefits, it may be reasonable for analysts to use CEA to avoid the burden of conducting a CBA. The third constraint arises when analysts deal with intermediate goods whose linkage to preferences is not clear. For example, the exact contribution of different types of weapon systems to overall national defense is often unclear. In such situations, CBA is not possible, but CEA may give useful information concerning the relative efficiency of alternatives.

Even though both CEA and CUA have been widely used by analysts working in a number of policy areas, especially health and defense, they are not applied consistently. There is great variation in a number of practices, such as the specification of alternative policies, the justification of effectiveness measures, the inclusiveness of cost measures, and the discounting of effectiveness over time. Indeed, one recent review of the application of CEA and CUA to the evaluation of health policies found that approximately half of the studies examined were of less than adequate quality.
CEA compares (usually mutually exclusive) alternatives on the basis of their costs and a single quantified but not monetized effectiveness measure, such as number of lives saved per dollar. Though there is no conceptual reason why costs cannot be measured comprehensively, in practice analysts most often measure them narrowly as budgetary costs. For example, CEA might include hospital salaries and supplies but might not include clients' waiting time. Except where otherwise noted, our discussion in this chapter refers to CEA as it is commonly practiced, that is, with this usually overly narrow specification of cost.

If budgetary cost happens to equal opportunity cost exactly, and the effectiveness measure is the only impact for which people are willing to pay, and the scale of the alternatives being compared is the same, then the rankings of alternatives by CEA and CBA will be identical. As discussed in Chapters 1 and 2, however, CBA not only produces a ranking of alternatives, it also reveals whether the highest-ranked or any of the other alternatives increase efficiency. CEA produces a ranking but does not provide explicit information about whether there would be positive net social benefits associated with any of the alternatives being considered. However, if all alternatives are mutually exclusive, and the status quo is among the alternatives, sharing similar scale and patterns of costs and benefits, then CEA does select the most efficient policy.

In many situations, the effectiveness measure selected by analysts (or decision makers) for use in CEA does not correspond to social benefits as measured in CBA, which are ultimately based on the willingness-to-pay (WTP) of individuals. We can reasonably infer in many circumstances that individuals would demonstrate WTP for incremental units of “effectiveness” such as “lives saved.” In other circumstances, however, the inference of WTP is more tenuous. For example, “number of addicts treated” may or may not be an approximate measure for such benefits as reductions in street crime and the other negative externalities of drug abuse. While analysts cannot avoid making estimates of WTP in doing CBA, even if they must rely on shadow prices from secondary sources, they often do not make an explicit connection between WTP and the effectiveness measure used in CEA. To highlight this problem, some authors distinguish between intermediate outputs, such as “patients appropriately treated,” where the value may not be clear, and final outputs, such as “lives saved,” for which people are more clearly willing to pay. Michael Drummond and his colleagues argue that “[i]ntermediate outputs are admissible, although care must be taken to establish a link between these and a final health output, or to show that the intermediate outputs themselves have some value... In general, though, one should choose an effectiveness measure relating to a final output.”

Cost-utility analysis also relates budgetary costs to a single benefit measure, but its benefit measure is a construct made up of several (usually two) benefit categories. For example, the benefit measure may be quality-adjusted life-years. If analysts use either additional years of life per dollar cost, or a quantitative index of improved quality of life per dollar cost, they are doing CEA. The rationale for CUA is that both the number of additional years and the quality of life during those years are important benefit categories. Thus, if the alternatives under consideration have two related quantified (but not monetized) benefits, or a benefit with two distinct dimensions that can
be quantified, CUA comes a step closer than CEA to the full treatment of benefits provided by CBA through WTP.

Keep in mind that CEA and CUA do not necessarily take account of all social costs. Indeed, most CEA and CUA studies consider only budgetary costs and exclude other social costs. In many studies, it is also unclear whether budgetary costs are based on marginal costs (the appropriate measure) or on average costs (which might differ widely from marginal costs). When alternatives have different opportunity costs that fall outside of measured costs, CEA and CUA may yield rankings that differ from those that would result from CBA.

**COST-EFFECTIVENESS ANALYSIS**

There are two basic ways to create cost-effectiveness ratios. For decision-making purposes, there are two ways to impose constraints to facilitate comparison of policy alternatives involving projects with different scales. There are also adjustments that can be made to make CEA closer to CBA.

**The Two CEA Ratios**

As CEA does not monetize benefits, it inevitably involves two different metrics: cost in dollars and an effectiveness measure—for example, lives saved, tons of carbon monoxide reduced, or children vaccinated. Because one cannot add or subtract non-commensurable metrics, one cannot obtain a single measure of net social benefits from the two metrics. One can only compute the ratio of the two measures as a basis for ranking alternative policies. Obviously, this can be done in two ways.

First, one can measure cost-effectiveness in terms of cost per unit of outcome effectiveness, for example, cost per life saved. To compute this, one takes the ratio of the budgetary cost of each alternative \(i\), denoted by \(C_i\), to the effectiveness (or benefit) of that alternative, \(E_i\):

\[
CE_i = \frac{C_i}{E_i}
\]

(13.1)

This CE ratio can be thought of as the average cost per unit of effectiveness. The most cost-effective project has the lowest average cost per unit of effectiveness. Thus, projects should be rank ordered from the most cost-effective (those with the smallest CE ratio) to the least cost-effective (those with the largest CE ratio).

Second, cost-effectiveness can be calculated as the ratio of the outcome effectiveness units per unit of budgetary cost, or:

\[
EC_i = \frac{E_i}{C_i}
\]

(13.2)

It is important to be aware that, rather confusingly, some authors call this EC ratio the cost-effectiveness ratio. This EC ratio can be thought of as the average effectiveness per unit of cost. The most cost-effective project has the highest average effectiveness per unit of cost. Thus, projects should be rank ordered from the most cost-effective (those with the largest EC number) to the least cost-effective (those with the smallest EC number).
Both cost-effectiveness measures involve computing for each alternative the ratio of the input to the output. Thus, they are measures of technical efficiency. As we discuss in the sections that follow, differences across policy alternatives in terms of scales of project, as well as the fact that cost-effectiveness measures often omit important social costs and benefits, frequently make them poor measures of allocative efficiency.

**CEA Where Scale Problems Are Irrelevant: Identical Program Budgets or Identical Program Effectiveness**

One may feel uneasy about selecting policy alternatives on the basis of their cost-effectiveness ratios. This intuition is correct. Ratios do not take into account the different scales of projects, a reason discussed in Chapter 2 for avoiding benefit-cost ratios.

If, however, all of the policy alternatives have the same cost, then there is no scale difference. If, in addition, the cost-effectiveness ratio is inclusive of all social costs and benefits, then CEA does rank alternatives in terms of allocative efficiency. Table 13.1 compares three alternative projects (one of which might be the status quo) for saving lives. The only (measured) costs are budgetary costs (in millions of dollars) and the effectiveness criterion is the number of lives saved. In this case the CE ratio reveals the average cost per life saved. Of course, in this simple example one does not even need to compute cost-effectiveness ratios: by “eye-ball” the table, one can easily observe that alternative C saves the most lives. Computing the cost-effectiveness ratio simply confirms this. It does not matter whether the ratio is calculated as cost per life saved or as lives saved per (million) dollars. Because all alternatives involve the same level of expenditure, they can be thought of as different ways of spending a fixed budget.

Similarly, scale is not a problem if the level of effectiveness is constant across all alternatives. This is illustrated in Table 13.2, which shows three alternatives for saving the same number of lives, namely, 10. Here, alternative A is best. Again, it does not matter whether the ratio is calculated as cost per life saved or as lives saved per (million) dollars. Situations in which the level of effectiveness is constant across alternatives, or is treated as constant, can be thought of as different ways of achieving a fixed effectiveness.

**Table 13.1 Cost-Effectiveness Analysis with Fixed (Identical) Costs**

<table>
<thead>
<tr>
<th>Cost and Effectiveness</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost Measure (budget cost)</td>
<td>$10M</td>
<td>$10M</td>
<td>$10M</td>
</tr>
<tr>
<td>Effectiveness Measure (number of lives saved)</td>
<td>5</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>CE Ratio (cost per life saved)</td>
<td>$2.0M</td>
<td>$1.0M</td>
<td>$0.67M*</td>
</tr>
<tr>
<td>EC Ratio (lives saved per million dollars)</td>
<td>0.5 life</td>
<td>1.0 life</td>
<td>1.5 lives*</td>
</tr>
</tbody>
</table>

*CE ratio or EC ratio of the most cost-effective alternative
TABLE 13.2 COST-EFFECTIVENESS ANALYSIS WITH FIXED (IDENTICAL) EFFECTIVENESS LEVELS

<table>
<thead>
<tr>
<th>Cost and Effectiveness</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost Measure (budget cost)</td>
<td>$5M</td>
<td>$10M</td>
<td>$15M</td>
</tr>
<tr>
<td>Effectiveness Measure (number of lives saved)</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>CE Ratio (cost per life saved)</td>
<td>$0.5M</td>
<td>$1.0M</td>
<td>$1.5M</td>
</tr>
<tr>
<td>EC Ratio (lives saved per million dollars)</td>
<td>2 lives</td>
<td>1 life</td>
<td>0.66 life</td>
</tr>
</tbody>
</table>

*CE ratio or EC ratio of the most cost-effective alternative

Note that in the case of fixed effectiveness, CEA corresponds to a simple cost-minimization problem (minimize dollars), while in the fixed-budget case CEA corresponds to a simple effectiveness-maximization problem (maximize lives saved). Both tables contain examples of dominated alternatives—by holding one dimension constant, they ensure that the alternative with the best cost-effectiveness ratio dominates on one dimension and is exactly the same on the other dimension. It is possible that one alternative can dominate another even if they have neither the same cost nor the same effectiveness, as long as it is superior on both dimensions. Clearly, dominated alternatives should not be selected. If an alternative dominates all others, then it should be selected.

Modifying CEA to Deal with Scale Differences

Large scale differences among alternatives potentially distort choice. The simple example in Table 13.3 illustrates this. It shows a choice between two mutually exclusive alternatives, A and B. Clearly, if we used a cost-effectiveness ratio, then we would choose alternative A. Yet, if we look more closely at alternative B, we see that it would save a large number of lives at the relatively low “price” per life saved of $0.5 million per life—much less than the shadow prices reviewed in Chapter 12. It is therefore likely that a CBA would show alternative B has larger net benefits. (Of course, we cannot be certain of this without valuing all costs and benefits.) Given that CEA was probably proposed in the first place because analysts were unwilling to monetize lives saved, how can CEA be used sensibly as a decision rule without monetizing lives saved?

TABLE 13.3 THE PROBLEM WITH THE CE RATIO WHEN SCALE DIFFERS

<table>
<thead>
<tr>
<th>Cost and Effectiveness</th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost Measure (budget cost)</td>
<td>$1M</td>
<td>$100M</td>
</tr>
<tr>
<td>Effectiveness Measure (number of lives saved)</td>
<td>4</td>
<td>200</td>
</tr>
<tr>
<td>CE Ratio (cost per life saved)</td>
<td>$250,000</td>
<td>$500,000</td>
</tr>
<tr>
<td>EC Ratio (lives saved per million dollars)</td>
<td>4.0 lives</td>
<td>2.0 lives</td>
</tr>
</tbody>
</table>

*CE ratio or EC ratio of the most cost-effective alternative
In order to make CEA more useful for decision making, decision makers sometimes specify a minimum acceptable level of effectiveness, denoted $E$. There are two common ways of imposing such a constraint.

First, we could select the project that meets the constraint at the lowest cost:

$$\text{Minimize } C_i$$
$$\text{s.t. } E_i > E\text{.}$$

Here the analyst or decision maker has decided on a minimum level of effectiveness and selects the least costly alternative to achieve it. The decision maker is acting as if he or she does not value additional units of effectiveness. This might apply, for example, to alternative ways of ensuring that children receive minimum amounts of fluoride to protect their teeth. It might also apply to some national defense activities, although in these examples, additional units of effectiveness above $E$ are probably worth something to decision makers.

Second, we could select the most cost-effective alternative that satisfies the effectiveness constraint:

$$\text{Minimize } CE_i$$
$$\text{s.t. } E_i > E\text{.}$$

This rule generally leads to higher levels of effectiveness and higher costs than the first rule.

An alternative way to constrain choice is to specify a maximum budgetary cost, denoted $\tilde{C}$. Again there are two alternative decision rules for selecting the best project subject to this constraint.

First, we could select the project that yields the largest number of units of effectiveness, subject to the budget constraint:

$$\text{Maximize } E_i$$
$$\text{s.t. } C_i < \tilde{C}\text{.}$$

The problem with this approach is that it ignores incremental cost savings. In other words, cost savings beyond $\tilde{C}$ are not valued.

Second, we could select the alternative project that most cost-effectively meets the imposed budget constraint:

$$\text{Minimize } CE_i$$
$$\text{s.t. } C_i < \tilde{C}\text{.}$$

This rule places some weight on incremental cost savings and is more likely to result in the selection of a project with less than the minimum cost.

**An Illustration of the Different CE Rules**

Imagine that each of the ten mutually exclusive and exhaustive projects shown in Table 13.4 are intended to save lives. The expected number of lives saved for each project are given in column 2, and the expected budgetary cost in millions of dollars.
for each project is in column 3. The "basic" cost-effectiveness ratio (cost per life saved) appears in column 4. Using the standard CE formula, projects can be ranked from most cost-effective to least cost-effective: project E is most cost-effective, followed by B, J, I, A, C, G, H, D, and F. Dominated projects can be eliminated from the choice set at the outset to simplify the analysis: Project D can be eliminated because it is dominated by project C, and projects C and F can be eliminated because they are dominated by project A. The most cost-effective alternative is project E. For this project, the average cost of a life saved is $2.0 million.

Project E, however, saves the fewest lives. Project B saves twice as many lives as project E and costs only $24 million more. Which project is better? This question illustrates the problem of different scales. Preferably, we would like the option of performing 2.2 project E's. This would be superior to project B, but it is not feasible because the projects are mutually exclusive and exhaustive.

This example illustrates that if we are prepared to monetize the value of a life saved, as in CBA, we can get closer to determining which alternative is the most allocatively efficient. Specifically, if a life saved is valued at more than $2.4 million, then project B is preferred to project E; if a life saved is valued at between $2.0 million and $2.4 million, then project E is preferred to project B; on the other hand, if a life saved is valued at less than $2.0 million, then no project at all is preferred to either project E or project B.

Now, suppose that the decision maker specifies that he or she wishes to save a minimum of 50 lives. The cheapest acceptable alternative is project H, but the most

<table>
<thead>
<tr>
<th>Projects</th>
<th>Lives Saved</th>
<th>Budget Cost ($M)</th>
<th>CE Ratio (Cost per life saved)</th>
<th>Budget Cost of Projects that Save at Least 50 Lives</th>
<th>CE Ratio of Projects that Save at Least 50 Lives</th>
<th>Lives Saved That Cost No More Than $250M</th>
<th>CE Ratio of Projects That Cost No More Than $250M</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>100</td>
<td>250</td>
<td>2.5</td>
<td>250</td>
<td>2.5*</td>
<td>100*</td>
<td>2.5</td>
</tr>
<tr>
<td>B</td>
<td>20</td>
<td>44</td>
<td>2.2</td>
<td>—</td>
<td>—</td>
<td>20</td>
<td>2.2</td>
</tr>
<tr>
<td>C</td>
<td>100</td>
<td>300</td>
<td>3.0</td>
<td>300</td>
<td>3.0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>D</td>
<td>50</td>
<td>300</td>
<td>6.0</td>
<td>300</td>
<td>6.0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>E</td>
<td>10</td>
<td>20</td>
<td>2.0*</td>
<td>—</td>
<td>—</td>
<td>10</td>
<td>2.0*</td>
</tr>
<tr>
<td>F</td>
<td>100</td>
<td>900</td>
<td>9.0</td>
<td>900</td>
<td>9.0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>G</td>
<td>60</td>
<td>210</td>
<td>3.5</td>
<td>210</td>
<td>3.5</td>
<td>60</td>
<td>3.5</td>
</tr>
<tr>
<td>H</td>
<td>50</td>
<td>200</td>
<td>4.0</td>
<td>200*</td>
<td>4.0</td>
<td>50</td>
<td>4.0</td>
</tr>
<tr>
<td>I</td>
<td>40</td>
<td>100</td>
<td>2.5</td>
<td>—</td>
<td>—</td>
<td>40</td>
<td>2.5</td>
</tr>
<tr>
<td>J</td>
<td>45</td>
<td>110</td>
<td>2.4</td>
<td>—</td>
<td>—</td>
<td>45</td>
<td>2.4</td>
</tr>
</tbody>
</table>

* CE ratio, budget cost, or effectiveness of the most efficient alternative.
cost-effective acceptable alternative is project A. Which is preferable? Note that project A costs $50 million more than project H, but it saves 50 more lives. The cost of these extra lives saved is only $1 million per life, on average. Saving these additional lives is more cost-effective than even project E, but it is 25 percent more expensive than project H. The choice depends on the decision maker’s willingness to trade additional lives saved for additional budgetary cost. Thus, even though CEA is often proposed as a way of avoiding monetization of some benefit, analysts or decision makers must often make trade-offs between costs and a nonmonetized benefit in order to make decisions.

The same type of problem arises if a budget constraint is imposed. Now the analyst should select either the project that yields the greatest benefit subject to the cost constraint or the most cost-effective project that satisfies the cost constraint. If the decision maker specifies a maximum budgetary cost of $250 million, project A saves the most lives, but project E is the most cost-effective. Again, to choose between projects A and E, the decision maker must consider trade-offs between additional lives saved and additional budgetary costs.

**Technical versus Allocative Efficiency: Omitted Costs and Benefits**

CEA almost invariably omits impacts that would be included in CBA. Indeed, CEA typically considers only one measure of effectiveness. Projects often have multiple benefits, however. For example, regulations that save lives may also reduce injuries or illnesses. On the cost side, as we have stressed previously, most CEA studies consider only budgetary costs. Relevant nonbudgetary opportunity costs may be omitted. To get a better measure of allocative efficiency, these costs and benefits should be taken into consideration. One way to get closer to doing this—that is, to reach a “halfway house” between standard CEA and CBA—is to compute the following ratio:

\[
\text{CE} = \frac{\text{social costs} - \text{other social benefits}}{\text{effectiveness}}
\]

If the numerator can be fully valued and monetized, then this adjusted CE ratio incorporates all the impacts that would be included in a CBA.

Most likely, however, CEA was selected in the first place because some social costs and benefits could not be monetized. Obviously, the omission of a particular category of social cost or benefit from the numerator could very well alter the ranking of alternatives. The danger of obtaining an arbitrary ranking increases as alternatives become less similar in terms of the inputs they require and the impacts they produce. Moreover, the transparency of CEA is also reduced because cost no longer has a simple interpretation (budgetary dollars) and decision makers must rely on the judgment of analysts about what social costs and benefits to include. For these reasons, *moving all the way to CBA with extensive sensitivity analysis is often a better analytical strategy overall than expanding the scope of measured costs in CEA.*
COST-UTILITY ANALYSIS

The greatest use of cost-utility analysis occurs in the evaluation of health policies. In CUA the (incremental) costs of alternative policies are compared to the health changes, usually measured in quality-adjusted life-years (QALYs), that they produce. CUA is most useful when a trade-off must be made between quality of life (morbidity) and length of life (mortality). In principle, however, CUA could be used with any two distinct dimensions of health status. CUA can be thought of as a form of CEA employing a more complex effectiveness measure; all of the previous discussion about decision rules thus applies. The rationale for distinguishing CUA is that considerable analytical effort has gone into the specific issues relating to developing QALYs.

The Meaning of Life—Quality-Adjusted Life-Years, That Is!

As QALYs involve two distinct variables—quality and quantity—the analyst must designate how these variables are to be defined and combined. This is a problem in multiattribute decision making. Consider, for example, the effects of three mutually exclusive alternative prenatal programs. Under the status quo, no babies with a particular condition are born alive. Prenatal alternative A will result in five babies being born alive per year, but with permanent, serious disabilities. Prenatal alternative B will result in only two live births, but with only low levels of disability. Before we can compare the costs of these alternatives to their effectiveness, we first have to make quantity and quality commensurate.

The general form of the problem is shown in Table 13.5. The columns show additional years of life ranging from a low of $Y_1$ to a high of $Y_5$. The rows show health status ranging from the worst (health state $H_1$), to the best (health state $H_5$). For simplicity, assume that alternatives A and B and the status quo involve the same costs and that there is no uncertainty about the longevities and health status they will yield. Suppose that the status quo (denoted SQ) gives $Y_{1H_1}$ (the fewest years of life in the worst health status), while alternative A achieves $Y_{2H_1}$ and alternative B achieves $Y_{3H_2}$. Clearly, the status quo is dominated, but how should we choose between alternatives A and B? Before answering this question, we must look more closely at the definition of health status.

<table>
<thead>
<tr>
<th>Health Status ($H$)</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$Y_4$</th>
<th>$Y_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$</td>
<td>$Y_{1H_1}^{SQ}$</td>
<td>$Y_{2H_1}$</td>
<td>$Y_{3H_1}$</td>
<td>$Y_{4H_1}$</td>
<td>$Y_{5H_1}$</td>
</tr>
<tr>
<td>$H_2$</td>
<td>$Y_{1H_2}$</td>
<td>$Y_{2H_2}$</td>
<td>$Y_{3H_2}$</td>
<td>$Y_{4H_2}^{B}$</td>
<td>$Y_{5H_2}$</td>
</tr>
<tr>
<td>$H_3$</td>
<td>$Y_{1H_3}$</td>
<td>$Y_{2H_3}$</td>
<td>$Y_{3H_3}^{A}$</td>
<td>$Y_{4H_3}$</td>
<td>$Y_{5H_3}$</td>
</tr>
<tr>
<td>$H_4$</td>
<td>$Y_{1H_4}$</td>
<td>$Y_{2H_4}$</td>
<td>$Y_{3H_4}$</td>
<td>$Y_{4H_4}$</td>
<td>$Y_{5H_4}$</td>
</tr>
<tr>
<td>$H_5$</td>
<td>$Y_{1H_5}$</td>
<td>$Y_{2H_5}$</td>
<td>$Y_{3H_5}$</td>
<td>$Y_{4H_5}$</td>
<td>$Y_{5H_5}$</td>
</tr>
</tbody>
</table>
Defining States of Health Status

Defining health status is complex. Health states are normally defined by CUA researchers in collaboration with clinicians familiar with variations in health—whether in relationship to particular diseases, injuries, and mental states, or to health in general. This reliance on experts is based on the assumption that neither the public nor potential treatment subjects are likely to have enough information and knowledge to formulate health states. Most often experts formulate health status indexes for specific diseases or illnesses.

George Torrance and his colleagues, however, have developed a comprehensive four-dimensional classification system with the following dimensions: physical function (mobility and physical activity); role function (ability to care for oneself); social-emotional function (emotional well-being and social activity), and "health problem" (including physical deformity).

Paul Kind and his colleagues have developed a disability ranking based on two dimensions: disability level and the level of distress. Their disability levels are: no disability, slight social disability, severe social disability or slight work performance impairment, choice of work or work performance seriously limited, unable to work or continue education, confined to chair or wheelchair, confined to bed, and unconscious. Their distress levels are: none, mild, moderate, and severe.

Formulating a Health Status Index

How are different health states scaled to form an index? How are changes in the index traded against additional years of life? Obviously, the usefulness of CUA depends on the validity of the methods used to answer these two questions. In the CUA literature, efforts to answer these questions are referred to as measuring the utilities, or utility values, of a sample of individuals.

There are three common methods of deriving utilities of health status: the health rating method, the time trade-off method, and the standard gamble method. The methods vary in the extent to which they correspond to the economic concept of utility.

Health rating method. Generally, analysts derive health rating from questionnaires or interviews with health experts or potential subjects of treatment, members of society in general, or on the basis of their own expertise. Respondents are presented a scale with well-defined extremes. For example, the scale may assign "death" a value of 0 and "health" a value of 1. Intermediate health states are described in detail to the respondents, who are then asked to locate each state between the end points, 0 and 1. If there are three intermediate health states described to an individual corresponding to "seriously disabled," "moderately disabled," and "minimally disabled," an individual might, for example, assign values of 0.15, 0.47, and 0.92 to these states, respectively.

This rating scale concerns only the health state dimension in Table 13.5. It does not directly provide a method for trading off health states with additional years. However, assuming that health status and longevity have independent effects on utility, the scale values can be directly merged with years of life to get QALYs, which can
then be used as an effectiveness measure. The two other methods we review scale the
health index and can analyze the trade-offs between the health states and years.

The time trade-off method. In the time trade-off method, respondents are
asked to compare different combinations of length and quality of life. The typical
comparison is between a longer life of lower health status and a shorter life with a
higher health status. Figure 13.1 illustrates such a comparison. The horizontal axis
measures additional years of life ($Y$) and the vertical axis measures health status ($H$).
Respondents might be asked to compare some status quo point, say $R$, representing
health status $H_2$ and additional years of life $Y_1$, with an alternative point, say $S$, repre-
senting health status $H_1$ and additional years of life $Y_2$. If a respondent is indifferent
between the two points, then he or she is willing to give up $H_2 - H_1$ units of health qual-
ity in return for $Y_2 - Y_1$ additional years of life. This pattern of trade-offs can provide a
basis for rank ordering the cells in Table 13.5.

This method assumes implicitly that additional years of life are valued equally,
that is, there is no discounting of health years.

The standard gamble method. In the standard gamble approach, respondents
are presented with a decision tree along the lines described in Chapter 6. Respondents
are offered a choice between two alternatives. Alternative $C$ has two possible out-
comes: either a return to normal health for $N$ additional years (occurring with proba-
bility $p$) or immediate death (occurring with probability $1 - p$). Alternative $C$ might be
an operation that has probability $1 - p$ of failure (death), but which, if successful, will

Figure 13.1  Time Trade-Off Example
return the patient to normal health for $N$ years. Alternative D guarantees the patient $t$ additional years with a specified level of health impairment. This choice is shown in Figure 13.2. The probability $p$ is varied until a respondent is indifferent between alternatives C and D. When using a health status index ranging from 0 (death) to 1 (normal health for $N$ years), the $p$ at which a respondent is indifferent can be interpreted as that respondent’s utility from alternative D.

The issues relating to QALYs are not all resolved. One problem is the issue of discounting additional years of life (the $Y$ axis in Table 13.5). The basic idea that individuals have positive discount rates relating to additional years is widely accepted, but there is considerable controversy over the theory, measurement, and level of the appropriate discount rate. We touched briefly on some of these issues in Chapter 5. One obvious problem when thinking about discounting health is that individuals cannot trade near (far) health years for far (near) health years. While many of the issues remain unresolved, it is problematic to discount costs but not to discount health years. The reason is that the cost-effectiveness ratio would improve if we delayed the health expenditure until the following year.

### EXHIBIT 13.1

Barbara McNeil and her colleagues conducted a study of laryngeal cancer treatment. First, the authors constructed a health status index for respondents in which 25 years of additional life (the full life expectancy for a 50-year-old male) with normal speech was assigned a value of 100 and immediate death was assigned a value of 0. Second, respondents were offered a choice between (1) a gamble with a 50 percent chance of survival for 25 additional years and a 50 percent chance of death in a few months (which we value as zero), and (2) a specified certain number of years of impaired survival ($t$) ranging between 0 and 25 years. The number of certain years offered with impairment the lower
branch in Figure 13.2) was varied until an individual was indifferent between the two branches. If an individual was indifferent when the value of the lower branch was seven years, for example, then seven additional years of certain life with impairment was assigned a value of 50 on the health status index. Additional gambles were offered to determine the periods of certain survival with impairment that would be assigned health status indexes of 25 and 75.

One of their findings was that executives and firefighters (males, with an average age of 40 years) would trade off 14 percent of their life expectancy to avoid artificial speech. Although most respondents would accept some decrease in life-years to avoid artificial speech, virtually none would accept any decrease in years when only five years or less of certain survival were offered.


Some Caveats on CUA

As these preference elicitation methods are based on questionnaires, many of the issues discussed in Chapter 11 with respect to contingent valuation are also relevant to QALY methods. For example, samples drawn from the general public tend to be more prone to the hypotheticality problems described in Chapter 11—respondents simply have not thought very much about these kinds of issues. This problem can be ameliorated somewhat by providing detailed descriptions of the various health states and the expected changes in years of life. Additionally, in practice, convenience samples rather than random samples are often drawn from the general population. Unfortunately, respondents who are potential treatment candidates because they are already ill may have incentives to exaggerate the “utility” they would receive from better health states. Finally, of course, either type of sample may be subject to framing (and neutrality) effects inherent in the questionnaire approach. Therefore, a utility elicitation instrument should be subjected to the same kind of validity checks as a contingent valuation instrument. In spite of all these caveats, there is evidence that different samples—whether the general public, potential treatment candidates, or physicians—as well as different sexes, ages, and races, do not vary much in their responses, either in terms of ranking health states or making trade-offs.15

A second type of problem arises from the fact that different utility derivation techniques produce different weightings. One recent survey, for example, found that six different methods to assess preferences did not correlate well.16

THE USE OF LEAGUE TABLES

CEA and CUA usually compare mutually exclusive projects. By definition, this means that the alternative projects address the same problems—for example, alternative methods of breast cancer screening. Yet, both CEA and CUA have been used to make rankings across policies that are not mutually exclusive. League tables draw on multiple CEA or CUA studies to rank policies sharing the same cost-effectiveness measure, usu-
ally "cost per life saved." John Morrall III, for example, has developed a league table of regulations intended to save lives. He found that regulations ranged in their cost-effectiveness from a cost per life saved of $100,000 (steering column protection regulation) to a cost per life saved of $132 million (cattle feed regulation). CUA league tables are usually used to "compare" investments in different kinds of health treatments.

How useful are league tables? Comparisons of mutually exclusive projects inherently control for some of the differences in the measurement of cost and effectiveness. There can be no such presumption when comparing studies across different authors, using different data, and somewhat different methodologies. Different studies may measure costs differently, they may omit different costs, some may include other social benefits, and they may differ considerably in scale. These problems apply even more so to CUA studies because different methodologies are used to calculate QALYs and, as discussed earlier, different QALY methodologies do not necessarily produce consistent results.

Karen Gerard's caution on the use of league tables is appropriate:

[A]s more and more studies were read in the course of the investigation, it became striking how many not only placed their results in some standard QALY league table, but also purported to present their results as 'favourable.' It is unlikely to be the case that all of these studies can have 'favourable' results.

Thus, caution is warranted in using league tables as guides for policy choice.

**CONCLUSION: WHEN IS CEA CLOSE TO CBA?**

CEA is closest to CBA when budgetary costs approximate social costs, when the effectiveness measure "captures" most of the social benefits, and when alternative projects are of similar scales. When there are significant nonbudgetary social costs or significant other categories of benefits, CEA is not close to CBA. When they are close, CEA may be less expensive and more transparent than CBA. When they are not close, analysts have three options. First, they should, if they can, do CBA. Second, they can move to a more qualitative evaluation method. Third, they can try to incorporate significant nonbudgetary social costs and other categories of benefits into cost-effectiveness measures.

**EXERCISES FOR CHAPTER 13**

1. A public health department is considering five alternative programs to encourage parents to have their preschool children vaccinated against a communicable disease. The following table shows the cost and number of vaccinations predicted for each program:

<table>
<thead>
<tr>
<th>Program</th>
<th>Cost ($)</th>
<th>Number of Vaccinations</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20,000</td>
<td>2,000</td>
</tr>
<tr>
<td>B</td>
<td>44,000</td>
<td>4,000</td>
</tr>
<tr>
<td>C</td>
<td>72,000</td>
<td>6,000</td>
</tr>
<tr>
<td>D</td>
<td>112,000</td>
<td>8,000</td>
</tr>
<tr>
<td>E</td>
<td>150,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>
a. Ignoring issues of scale, which program is most cost-effective?
b. Assuming that the public health department wishes to vaccinate at least 5,000 children, which program is most cost-effective?
c. If the health department believes that each vaccination provides social benefits equal to $20, then which program should it adopt?

2. Analysts wish to evaluate alternative surgical procedures for spinal cord injuries. The procedures have various probabilities of yielding the following results:

- Full recovery (FR)—the patient regains full mobility and suffers no chronic pain.
- Full functional recovery (FFR)—the patient regains full mobility but suffers chronic pain that will make it uncomfortable to sit for periods of longer than about an hour and will interfere with sleeping two nights per week, on average.
- Partial functional recovery (PFR)—the patient regains only restricted movement that will limit mobility to slow-paced walking and will make it difficult to lift objects weighing more than a few pounds. Chronic pain is similar to that suffered under full functional recovery.
- Paraplegia (P)—the patient completely loses use of legs and would, therefore, require a wheelchair or other prosthetic for mobility, and suffers chronic pain that interferes with sleeping four nights per week, on average. Aside from loss of the use of his or her legs, the patient would regain control of other lower body functions.

a. Describe how you would construct a quality-of-life index for these surgical outcomes by offering gambles to respondents. Test your procedure on a classmate, friend, or other willing person.

b. Assume that the index you construct on the basis of your sample of one respondent is representative of the population of patients. Use the index to measure the effectiveness of each of three alternative surgical procedures with the following distributions of outcomes:

<table>
<thead>
<tr>
<th>Surgical Procedures</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR</td>
<td>.10</td>
<td>.50</td>
<td>.40</td>
</tr>
<tr>
<td>FFR</td>
<td>.70</td>
<td>.20</td>
<td>.45</td>
</tr>
<tr>
<td>PFR</td>
<td>.15</td>
<td>.20</td>
<td>.10</td>
</tr>
<tr>
<td>P</td>
<td>.05</td>
<td>.10</td>
<td>.05</td>
</tr>
</tbody>
</table>

c. Imagine that the surgical procedures involved different life expectancies for the various outcomes. Discuss how you might revise your measure of effectiveness to take account of these differences.

NOTES

1 For a discussion of other less comprehensive techniques, such as cost minimization, see Michael Drummond, Greg L. Stoddart, and George W. Torrance, Methods for the Economic Evaluations of Health Care Programmes (Oxford: Oxford University, 1987), pp. 7–9 and 39–73.


3 Drummond, Stoddart, and Torrance, Methods for the Economic Evaluations of Health Care Programmes, p. 76.
Gerard found that approximately 90 percent of the health studies she reviewed only looked at direct budgetary costs, "Cost-Utility in Practice," p. 263.

3Ibid., pp. 263–264.

4See, for example, Karin V. Lowson, M.F. Drummond, and J.M. Bishop, "Costing New Services: Long-Term Domiciliary Oxygen Therapy," The Lancet, no. 8230 (May 1981), 1146–1149.

5Dominant alternatives tend to be rare but they can occur. For an example, see Russell D. Hall, Jack Hirsh, David L. Sackett, and Greg L. Stoddart, "Cost-Effectiveness of Primary and Secondary Prevention of Fatal Pulmonary Embolism in High-Risk Surgical Patients," Canadian Medical Association Journal, 127, no. 10 (November 13, 1982), 990-995.

6The term s.t. means "subject to."


10For an overview and discussion of this issue, see Debra Froberg and Robert L. Kane, "Methodology for Measuring Health-State Preferences - IV: Progress and a Research Agenda," Journal of Clinical Epidemiology, 42, no. 7 (1989), 675–685.


12Emmett B. Keeler and Shan Cretin, "Discounting of Life-Saving and Other Nonmonetary Effects," Management Science, 29, no. 3 (March 1983), 300–306. However, this is not a paradox per se because, as we have shown, a CE ratio never tells us whether a project has positive social value and hence, should be implemented—whether this year or next year.


16See, for example, Alan Williams, "Economics of Coronary Artery Bypass Grafting," British Medical Journal, 291, no. 6491 (August 3, 1985), 326–329.

Government policies, programs, and projects typically affect individuals differently. Thus, in conducting CBAs, analysts often report benefits and costs for separate categories of people. The relevant classification of individuals into groups for this purpose depends, of course, on the specific policy under evaluation. Some examples include: consumers versus producers versus taxpayers; program participants versus nonparticipants; citizens (of a nation or a state or a city) versus noncitizens; and high-income persons versus low-income persons.

Once individuals are divided into categories, the first issue that must be decided is whether each group will be given standing in the CBA. For example, in conducting a CBA of U.S. regulatory policy on acid rain, a decision must be made as to whether to give standing to Canadians affected by acid rain resulting from manufacturing in the United States.

Given this decision, costs and benefits may be reported separately for each group receiving standing. But how is this information to be utilized in making a decision concerning the policy that has to be analyzed?

Throughout this book, we have emphasized use of the Kaldor-Hicks potential compensation test in reaching such decisions. In using this test, benefits and costs are simply summed across all groups with standing to determine whether total benefits are larger than total costs and, hence, whether the policy should be adopted. This test examines benefits and costs from the perspective of society as a whole, where "society" is composed of all groups with standing. Indeed, the implicit philosophy behind the Kaldor-Hicks potential compensation test is that, given standing, it does not matter who receives the benefits from a government program or who pays the costs ("a dollar is a dollar regardless of who receives or pays it"), as long as there is a net gain.
to society as a whole—in other words, as long as the program is efficient in terms of potential Pareto improvement. Strict use of the Kaldor-Hicks test means that information on how benefits and costs are distributed among groups is ignored in decision making.

In making actual policy decisions, however, the way in which benefits and costs are distributed among various groups is seldom ignored. In fact, this consideration can have a major influence over whether a policy is politically acceptable. Hence, in actual decision making, a dollar received or expended by a member of one group may not be treated as equal to a dollar received or expended by a member of another group.

In this chapter, we focus on the role of the distribution of benefits and costs among groups in using CBA for decision making. We first examine the economic rationale for treating dollars received or expended by various groups differently in CBA. We then consider approaches for doing this in practice.

DISTRIBUTIONAL JUSTIFICATIONS FOR INCOME TRANSFER PROGRAMS

The rationale suggested by economists for treating dollars received or expended by various groups differently in CBA is mainly limited to situations in which low-income persons are helped (or hurt) by a program more than other persons. Political decision makers may, of course, treat dollars received or expended by various groups differently, even if their income levels are similar. They may, for example, be influenced by differences among groups in voting behavior or campaign contributions. Economists, however, typically argue for treating dollars received or expended by various groups similarly unless they differ in terms of income or wealth. Consequently, in the remainder of this chapter, we focus on CBAs of policies that have differential effects on groups that differ by income—for example, projects that are located in underdeveloped regions or programs that are targeted at disadvantaged persons.

To illustrate such a policy, consider a hypothetical program that taxes high-income persons in order to provide income transfers to low-income persons. The tax component of this program is illustrated in Figure 14.1. For purposes of discussion, assume that the market represented in this graph is for a luxury good, such as yachts, that is only purchased by the rich. In the absence of the tax, equilibrium in this market would occur at a price of $P_1$ and a quantity of $Q_1$. If an excise tax of $t$ is levied against each unit of output, then the supply curve would shift up by this amount as suppliers attempted to pass along to consumers the additional cost the tax imposes upon them.

Using the approach implied by the Kaldor-Hicks rule, a distributional analysis of the costs and benefits associated with this tax would look like this:

<table>
<thead>
<tr>
<th></th>
<th>BENEFITS</th>
<th>COSTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Producers</td>
<td>$A + C$</td>
<td>$B + D$</td>
</tr>
<tr>
<td>Consumers</td>
<td>$C + D$</td>
<td>$A + B$</td>
</tr>
<tr>
<td>Transfer Recipients</td>
<td>$A + B$</td>
<td></td>
</tr>
<tr>
<td>Society</td>
<td>$C + D$</td>
<td>$A + B$</td>
</tr>
</tbody>
</table>
Thus, a deadweight loss equal to areas $B + D$ would result from the tax. In addition to this deadweight loss, there are two off-graph social costs that also would result from our hypothetical transfer program:

1. Administering both the tax and the transfer parts of the program would require the use of social resources.
2. Some of those receiving the transfer would probably work less or stop working entirely, thereby reducing the goods and services available to society. In fact, there is considerable evidence that this is exactly what occurs under existing welfare programs. Only part of this loss would be offset by the gains in leisure to transfer recipients (see Chapter 9). The remaining residual is a second source of deadweight loss.

It should be obvious that this program would never pass the Kaldor-Hicks test. It must instead be justified on distributional grounds. In other words, one would have to argue that giving a low-income person a dollar warrants taking more than a dollar away from a higher-income person. Apparently, this distributional argument has some force because programs such as AFDC and food stamps, which transfer income from higher-income to lower-income persons do in fact exist. Hence, it appears that society is willing to sacrifice some efficiency in order to provide assistance to low-income persons.

The importance of this for CBA analysis is that it implies that, in practice, a dollar of benefits received or a dollar of costs incurred by a low-income person is sometimes
given greater weight in assessing government programs than a dollar of benefits received or a dollar of costs incurred by a higher-income person. How can this be justified?

THE CASE FOR TREATING LOW- AND HIGH-INCOME GROUPS DIFFERENTLY IN CBA

In the economics literature, there are at least three arguments for giving dollars received or paid by low-income persons greater weight in CBA than dollars received or paid by higher-income persons:

1. Income has diminishing marginal utility.
2. The income distribution should be more equal.
3. The "one man–one vote" principle should apply.

We discuss each of these arguments in turn.

Diminishing Marginal Utility of Income

The first argument is based on the standard assumption in economics that each additional dollar an individual receives provides less utility than the preceding dollar. A corollary of this assumption is that a dollar received or a dollar of cost incurred by a high-income person has less of an impact on his or her utility than it would on a low-income person's utility. For that reason, the argument suggests, it should count less in a CBA. This argument can be summarized algebraically as follows:

\[ \frac{\Delta u_l}{\Delta y_l} > \frac{\Delta u_h}{\Delta y_h} \]  

where \( \Delta u_l/\Delta y_l \) is the marginal private utility of income of individual \( l \) indicates a low-income person, and \( h \) a high-income person.

Income Distribution Should Be More Equal

The second argument for giving dollars received or paid by the poor greater weight in CBA than dollars received or paid by the rich is premised on the assertion that the current income distribution is less equal than it should be and social welfare would be higher if it were more equal. There are several possible bases for such an assertion. The first is that a highly unequal distribution of income may result in civil disorder, crime, and riots. More equality in income may reduce these threats to the general social welfare. Second, it can be argued that there is some minimum threshold of income that is so low that no one can (or, to preserve human dignity, should have to) live below it. This suggests that the income distribution be made more equal by truncating it at the minimum threshold through income floors. Third, some relatively well-off persons may receive utility if the circumstances facing the worse-off members of society at the bottom of the income distribution improve. Certain types of charitable giving, such as contributions to the Salvation Army, provide some evidence for the
existence of this form of altruism. Finally, it is possible that some persons value greater income equality in and of itself.9

If for any of these reasons society prefers greater income equality than currently exists, then a dollar increase in the income of a low-income person would result in a larger increase in the welfare of society as a whole than would a dollar increase in the income of a high-income person. Note that this would be true even if the marginal utility of income was not diminishing and, consequently, a dollar increase in the income of high- and low-income persons resulted in equal increases in the utilities of these persons. Each of the justifications listed in the previous paragraph suggests that society as a whole (or at least some relatively well-off members of society) becomes better off if those at the bottom of the income distribution gain relative to those in the rest of the distribution. Thus, the first and second arguments are quite distinct from one another.

Stated algebraically, the second argument implies that

\[ \Delta SW/\Delta y_i > \Delta SW/\Delta y_h, \text{ even if } \Delta u_i/\Delta y_i = \Delta u_h/\Delta y_h \]  

where \( \Delta SW \) refers to the change in aggregate social welfare and \( \Delta SW/\Delta y_i \) is the marginal effect on social welfare of a change in income that is received by individual \( i \).10

This argument confronts the Kaldor-Hicks test quite directly. It implies that there are some projects and programs that fail the Kaldor-Hicks test but should nonetheless be adopted if they redistribute income in a way that makes the income distribution more equal. In other words, the argument suggests that some programs that are inefficient should be undertaken if they increase income equality sufficiently. This also implies, of course, that some projects that make the income distribution less equal should not be undertaken, even though the Kaldor-Hicks test implies that they are efficient.

The “One Person–One Vote” Principle

This argument begins by noting that the benefits and costs of government programs to consumers are appropriately measured as changes in consumer surplus. Then it goes on to point out that because high-income persons have more income to spend than low-income persons, the measured impacts of policies on their consumer surplus will typically be larger and, hence, will be of greater consequence in a CBA based strictly on the Kaldor-Hicks rule.

This is illustrated by Figure 14.2, which compares the demand schedules of a typical high-income consumer and a typical low-income consumer for a good. If the good is a normal good, that is, if demand for the good increases as income increases, then the demand schedule of the high-income consumer will be to the right of that of the low-income consumer, as the diagram shows. Now if a government policy increases the price of the good, say from \( P_1 \) to \( P_2 \), both consumers will bear the cost of that increase in the form of a loss in consumer surplus. However, the loss suffered by the high-income consumer (areas \( A + B \)) will be greater than the loss borne by the low-income consumer (area \( A \) alone). As a result, a CBA will give more weight to the impact of the policy on the high-income consumer than on the low-income consumer.
The final part of the argument suggests that in a democracy, low-income persons should have as much influence over decisions on whether to undertake public projects as high-income persons. In other words, it is argued that "since the principle of 'one man, one vote' is deeply embedded in the concept of democracy," measures of changes in consumer surplus for different persons should be adjusted to what they would be if everyone had the same income. For example, we would count the impact of the price change on the low-income person represented in the diagram at about double what we would count the impact on the high-income person, in effect, equalizing the "votes" of the two individuals.

**DISTRIBUTIONAL WEIGHTS**

In principle, various groups can be treated differently in a CBA by using distributional weights. As will be seen, distributional weights are just numbers, such as 1, 2, or 1.5, which are intended to reflect the value placed on each dollar paid out or received by each group. For example, the highest-income group might be given a value of 1, while a lower-income group could be given a value of 2, implying that a dollar received by the member of the low-income group is valued in the CBA at twice that of a dollar received by the high-income group.

Distributional weights can be incorporated into a CBA through a slight modification of the net present value formula:

\[
\text{NPV} = \sum_{j=1}^{m} W_j \sum_{t=0}^{\infty} \frac{b_{j,t} - c_{j,t}}{(1+r)^t}
\]

(14.3)

where \(W_j\) is the distributional weight for group \(j\), \(b_{j,t}\) are the real benefits received by group \(j\) in period \(t\), \(c_{j,t}\) are the real costs imposed on group \(j\) in period \(t\), \(m\) is the number of groups, and \(r\) is the real social discount rate.
The idea behind this formula is simple. The persons affected by the government policy are divided into as many groups as seems appropriate. Each group is then given a distributional weight. The net present value for each group is then computed and multiplied by its weight. These weighted net present values are then added together to obtain an overall net present value. Note that in CBAs that rely strictly on the Kaldor-Hicks rule, \( W_j \) is implicitly set equal to 1 for all groups.

To illustrate the use of distributional weights, consider the following simple numerical example, a welfare program that costs the nonpoor $10 billion and raises the incomes of the poor by $8 billion:

\[
\begin{align*}
&b_{\text{poor}} = $8 \text{ billion} \\
&c_{\text{nonpoor}} = $10 \text{ billion}.
\end{align*}
\]

This program obviously would not pass the standard Kaldor-Hicks test. It is inefficient. But, imagine that distributional weights have been developed that imply that a dollar received by a poor person is worth 50 percent more than a dollar received by a nonpoor person. That is:

\[
W_{\text{poor}} = 1.5 \\
W_{\text{nonpoor}} = 1.
\]

In this case, the costs and the benefits of the program would be multiplied by the appropriate weights:

\[
(1.5)($8 \text{ billion}) - (1)($10 \text{ billion}) = +$2 \text{ billion}.
\]

Thus, with the weights, the program passes muster; but without them, it does not.

**DETERMINING DISTRIBUTIONAL WEIGHTS**

The obvious difficulty with implementing this approach is determining appropriate weights for each group. The weights should, of course, be consistent with the rationale for using them. However, of the three arguments for using weights presented earlier, only the one based on the one man–one vote principle suggests an approach that could potentially be used in practice for valuing the weights of various groups, and even in this case, the information requirements are substantial. Among the required information is the average income level of each relevant group, an estimate of the income elasticity of demand for each good affected by the government policy being evaluated (i.e., the percentage change in the quantity demanded of each good that results from a 1 percent increase in income), and an estimate of the market demand curve for each affected good. Given this information, the consumer surplus of the average member of each group could then be computed. These estimates could, in turn, be used to derive distributional weights for each group that are consistent with the one man–one vote principle.¹³

To develop weights that are consistent with the remaining two arguments, information is needed on \( \Delta u/\Delta y \) (the marginal private utility of income) and \( \Delta SW/\Delta y \) (the marginal effect on social welfare of a change in income) for a typical member of each group of interest. For example, a set of distributional weights could be developed that is consistent with the diminishing marginal utility of income argument if we knew that a $100 increase in income increased the utility of rich people by, say, two units and
the utility of poor people by three units by simply computing weights on the basis of the ratio of the marginal utility values. Thus, \( W \) would be set equal to 1 for rich persons and to 1.5 for poor persons. Similarly, a set of distributional weights that is consistent with the argument that the income distribution should be more equal could be developed if we knew that a $100 increase in the income of a typical poor person increased social welfare by two units, but the same increase in the income of a typical rich person increased social welfare by only one unit.

Unfortunately, such information is not available and there is no known way to obtain it. First, utility is a subjective concept that defies cardinal measurement. Indeed, most economists eschew any explicit interpersonal comparisons of utility for this reason. But in the absence of interpersonal comparisons of utility, it is not possible to develop a system of relative distributional weights that is consistent with the diminishing marginal utility of income argument. Second, there is no general consensus among members of society concerning the specific relationship between a given change in the income levels of individuals and social welfare, except that most persons would, perhaps, agree that the relation is positive and its magnitude is larger for low-income than high-income persons. Without such a consensus, however, it is not possible to develop distributional weights that are consistent with the greater income equality argument.

**POLITICALLY DETERMINED WEIGHTS**

Given these problems in deriving distributional weights, how can distributional issues be actually handled in CBA? One possibility is to use contingent valuation techniques to determine distributional weights. Distributional weights, after all, can be viewed as a type of shadow price, an effort to measure benefits and costs in terms of their true social value. To the best of our knowledge, however, no attempt has so far been made to do this. If such an attempt were made, it would, of course, be subject to the limitations of contingent valuation described in Chapter 11.

Another approach is to derive distributional weights on the basis of revealed political behavior. For example, Otto Eckstein suggested that marginal income tax rates might be used to determine weights for different income classes. For instance, families with annual incomes between $15,000 and $30,000 face a marginal tax rate of 15 percent under the U.S. income tax, while families with over $30,000 in income face a marginal tax rate of between 28 percent and 33 percent. One might infer that in establishing these tax rates society, or at least Congress, felt that at the margin taxing away 15 cents from a low-income person was as painful as taxing away around 30 cents from a higher-income person. Taking this one step further, one could infer that the political process has implied that dollars paid out or received by moderate-income families under a government program should be weighted twice as heavily as dollars paid out or received by families at the top of the income distribution.

Although weighting schemes based on tax rates have been used in CBA, such schemes raise serious issues. One problem is that tax rates do not only reflect what
Congress thinks the relative value of a dollar to different income groups should be. For example, Congress may be afraid to set tax rates too high because of concern that high rates will create work and investment disincentives. In addition, Congress may use the progressive federal income tax to offset the fact that other taxes, such as state sales taxes, tend to be regressive. Indeed, it might be better to base weighting schemes on the total tax system than on the federal income tax alone.

An alternative to using marginal tax rates to derive distributional weights is to use public expenditure decisions instead.\textsuperscript{18} To illustrate, consider two projects, A and B. If project A has a higher net present value than project B—that is, $\text{NPV}^A > \text{NPV}^B$—yet project A is rejected and B is undertaken, this suggests that decision makers must have viewed project B as at least as desirable as project A on nonefficiency grounds. One possible explanation is that project B provides greater net gains to low-income persons than does project A (that is, $\text{NPV}^A < \text{NPV}^B$), while higher-income persons would be better off under project A ($\text{NPV}^A > \text{NPV}^B$).

If these conditions hold and the values of $\text{NPV}^A$, $\text{NPV}^B$, $\text{NPV}_l^A$, and $\text{NPV}_h^B$ are known, then it is possible to derive distributional weights by solving the following two simultaneous equations for $W_l$ and $W_h$:\textsuperscript{19}

\[
W_l \cdot \text{NPV}_l^A + W_h \cdot \text{NPV}_h^A = \text{NPV}^A \\
W_l \cdot \text{NPV}_l^B + W_h \cdot \text{NPV}_h^B = \text{NPV}^A
\] (14.4)

The second equation is based on the premise that because decision makers selected project B over project A, even though $\text{NPV}^A$ was actually larger than $\text{NPV}^B$, they must have implicitly used a set of weights that allowed them to treat project B as if it had a net present value that was at least as high as $\text{NPV}^A$. The weights are derived by treating the two projects as if, once weighted, they were of exactly equal value in the eyes of the decision makers.

The estimates of $W_l$ and $W_h$ should provide a reasonable approximation of the implicit weights that decision makers used in deciding between the two projects. Unfortunately, however, the estimates may reflect other factors than only the concern of decision makers over the trade-off between efficiency and distributional equity. Considerations related to electoral prospects, for example, may very well have been more important in determining the decision.

**A PRAGMATIC APPROACH TO WEIGHTING**

Given the enormous practical problems with obtaining a defensible set of distributional weights, we suggest that their use be limited to only those CBAs where distributional issues are of central concern—for example, CBAs of programs targeted at disadvantaged groups or at impoverished areas within countries, states, or cities. Even then, it may often be possible to use an approach that highlights the importance of the distributional implications associated with the policy being analyzed without requiring that any particular set of distributional weights be selected as the "correct" set.
To illustrate the particular approach that we suggest using to do this, we return to the Work/Welfare Initiative Demonstrations that we first described in Chapter 9. As indicated there, these demonstrations were run in a randomized experimental setting during the 1980s to determine the impacts of various combinations of job search, training, and subsidized jobs for AFDC recipients. Because these experimental programs were targeted at welfare recipients—an especially disadvantaged low-income group—both their distributional effects and their effects on economic efficiency are relevant. Thus, CBAs of the Work/Welfare Initiative Demonstrations should take both types of effects into account.

Displaying Unweighted Cost and Benefit Estimates

The first step in doing this is simply to display unweighted program impacts on society as a whole, as well as on pertinent subgroups. This is accomplished in the first three columns in Table 14.1, which simply duplicate total net present value estimates for the Work/Welfare Initiative Demonstrations that were originally reported in Table 9.4.20 Column 1 reports these estimates from the perspective of program participants, column 2 from the perspective of nonparticipants, and column 3 (which is computed by summing the first two columns) from the perspective of society as a whole.

None of the estimates reported in the first three columns of Table 14.1 are weighted. Even when distributional weighting is used, unweighted estimates of benefits and costs for society as a whole should always be provided in CBAs. In addition, whenever distributional considerations are important, benefit and cost estimates for relevant subgroups should also be provided, if it is feasible to do so.

The unweighted net present value estimates reported in the first three columns of Table 14.1 have two especially important implications. First, they indicate that almost all the Work/Welfare Demonstration programs pass the Kaldor-Hicks test. Specifically, 16 of the 19 unweighted net present value estimates for society as a whole, which appear in column 3, are positive. Second, in 6 of the 16 cases that pass the Kaldor-Hicks test, program participants were apparently made worse off by the tested program, while nonparticipants—a group who, on average, enjoys substantially higher-incomes than participants—were made better off. In these instances, a trade-off occurs between economic efficiency and distributional considerations. (A similar trade-off would also arise if a program failed the Kaldor-Hicks test, but participants were made better off. However, this situation did not arise in any of the Work/Welfare Demonstrations.) It is only when such a trade-off occurs that distributional weighting is relevant.

Conducting Sensitivity Tests

Column 4 tests whether the estimates reported in column 3 of Table 14.1 are sensitive to the choice of distributional weights. Thus, in contrast to the unweighted figures appearing in column 3, which are based on the assumption that society values the gains and losses of AFDC recipients and nonrecipients equally, those appearing in column 4 assume that the gains and losses of AFDC recipients are valued by society at
TABLE 14.1 SENSITIVITY OF THE WORK/WELFARE INITIATIVE DEMONSTRATION ESTIMATES TO THE USE OF DISTRIBUTIONAL WEIGHTS

<table>
<thead>
<tr>
<th></th>
<th>Social Net Present Value if Participant Perspective of Internal Distributional Weight = 2 [2 \times \text{Col 1} + \text{Col 2}]</th>
<th>Estimates of Internal Weights for Participants [\text{Col 2}/\text{Col 1}]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Net Present Value From Participant Perspective</td>
<td>Net Present Value From Nonparticipant Perspective</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>AFDC-R</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>APPLICANTS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Diego EPP/EWEP:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job search only</td>
<td>$644</td>
<td>$452</td>
</tr>
<tr>
<td>Job search/CWEP</td>
<td>798</td>
<td>1,156</td>
</tr>
<tr>
<td>San Diego SWIM</td>
<td>(880)</td>
<td>1,633</td>
</tr>
<tr>
<td>Virginia</td>
<td>1,134</td>
<td>667</td>
</tr>
<tr>
<td>West Virginia</td>
<td>(481)</td>
<td>389</td>
</tr>
<tr>
<td><strong>RECIPIENTS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Diego SWIM</td>
<td>725</td>
<td>1,698</td>
</tr>
<tr>
<td>Virginia</td>
<td>574</td>
<td>190</td>
</tr>
<tr>
<td>West Virginia</td>
<td>80</td>
<td>873</td>
</tr>
<tr>
<td>New Jersey</td>
<td>1,262</td>
<td>1,069</td>
</tr>
<tr>
<td>Maine</td>
<td>3,182</td>
<td>(418)</td>
</tr>
<tr>
<td><strong>APPLICANTS AND RECIPIENTS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cook County:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job search only</td>
<td>(420)</td>
<td>475</td>
</tr>
<tr>
<td>Job search/CWEP</td>
<td>(34)</td>
<td>362</td>
</tr>
<tr>
<td>Baltimore</td>
<td>1,739</td>
<td>74</td>
</tr>
<tr>
<td>Arkansas</td>
<td>(449)</td>
<td>944</td>
</tr>
<tr>
<td><strong>AFDC-U</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>APPLICANTS ONLY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Diego EPP/EWEP:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job search only</td>
<td>(1,196)</td>
<td>1,229</td>
</tr>
<tr>
<td>Job search/CWEP</td>
<td>(1,443)</td>
<td>1,414</td>
</tr>
<tr>
<td>San Diego SWIM</td>
<td>543</td>
<td>1,577</td>
</tr>
<tr>
<td><strong>RECIPIENTS ONLY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Diego SWIM</td>
<td>(921)</td>
<td>2,487</td>
</tr>
<tr>
<td><strong>APPLICANTS AND RECIPIENTS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baltimore</td>
<td>(1,233)</td>
<td>(1,856)</td>
</tr>
</tbody>
</table>

NA: Not Applicable
Source for columns 1–3: Table 9.4.
twice those of nonrecipients. In other words, participants are given a distributional weight of 2 and nonparticipants a weight of 1. Although these weights are obviously arbitrary—as previously stressed, we do not know the actual relative values that society places on dollars received and paid out by participants and nonparticipants—in our judgment, it seems likely that they overstate society’s generosity toward AFDC recipients and thus test whether the net social gain and loss estimates are sensitive to a rather extreme assumption.

A comparison of columns 3 and 4 indicates that this assumption causes ten of these estimates to become larger and nine smaller. More importantly, however, only three change sign. Although all three of these sign changes are from positive to negative—that is, from a net gain to net loss—in two cases, the originally estimated net gain was well under $100. Hence, it appears that conclusions concerning whether most of the Work/Welfare Initiative Demonstrations listed in the table were cost-beneficial to society are not very sensitive to the choice of distributional weights.

**Computing Internal Weights**

Column 5 in Table 14.1 is based on the computation of *internal weights*, an alternative to the approach used in column 4 for distributional weighting. This scheme works best if there are just two pertinent groups, one of which is relatively disadvantaged (e.g., participants in the Work/Welfare Demonstrations) and the other relatively advantaged (e.g., nonparticipants in the demonstrations).

Internal distributional weights are derived by first setting the weight for the advantaged group equal to unity and then computing the weight for the disadvantaged group by dividing the estimated net present value for the advantaged group by the estimated net present value for the disadvantaged group. The idea is similar to that behind the computation of internal rates of return. Instead of somehow selecting weights, one finds the weights at which the program being analyzed would just break even, in other words, the weights at which the net present value for society as a whole would equal zero. Viewed a bit differently, the internal weight for the disadvantaged group indicates the dollars of costs incurred by the advantaged group per dollar of benefits received by the disadvantaged group, if the former is made worse off by the program and the latter better off, or the dollars of benefits received by the advantaged group per dollar of costs incurred by the disadvantaged group, if the former is made better off and the latter worse off.

In Table 14.1, we compute internal weights for Work/Welfare Demonstration participants by dividing column 2 by column 1. We do this, however, only when there is a trade-off between efficiency and distribution, that is, when column 1 and column 3 are of the opposite sign. In all other instances (i.e., when columns 1 and 3 are both positive or both negative), a trade-off between efficiency and distribution does not exist and, consequently, distributional weighting is not germane. As a trade-off between efficiency and distribution only occurred in six instances in the Work/Welfare Demonstrations, only six internal weights for program participants appear in column 5 of Table 14.1.
Each of these six values indicates the weight at which a demonstration program would just break even. Thus, if the “true” weight for participants is larger than their internal weight, programs with positive unweighted social net present values would fail to break even once their distributional implications were taken into account and programs with negative unweighted social net present values would more than break even. However, because the “true” weight for participants is unknown, policymakers would have to make a judgmental decision as to whether dollars of benefits or costs to participants should be given a higher or lower value than that implied by the computed internal weights. Indeed, one advantage of internal weighting is that it makes the trade-off between efficiency and distribution explicit for policymakers.

For example, in all six of the cases for which internal weights are computed in Table 14.1, program participants were worse off under the demonstration programs, but the unweighted social net present value estimate was positive. Thus, all six programs pass the Kaldor-Hicks test, even though they make the distribution of income less equal. If policymakers believed that dollars lost to participants in the Work/Welfare Demonstrations should be valued at, say, 25 percent more than dollars gained by nonparticipants, this would imply that two of these six programs failed to break even and, hence, should be discontinued even though they passed the Kaldor-Hicks test. If they instead valued dollars lost to participants at three times more than dollars gained by nonparticipants, then this would imply that five of the six programs failed to break even. Thus, they would conclude that five of the programs that passed the Kaldor-Hicks test actually had a negative payoff once their adverse effects on the income distribution were taken into account. But is a weight for participants as high as 3 plausible? We consider this issue next.

Obtaining Upper-Bound Values for Distributional Weights

It was pointed out earlier in this chapter that pure transfer programs inevitably fail the Kaldor-Hicks test—each dollar of transfer benefits costs nonrecipients more than a dollar. However, it has been argued that transfer programs can be used as a standard to which other types of programs that redistribute income can be compared. Specifically, the argument suggests that if a nontransfer program makes the disadvantaged better off, but results in a loss of efficiency, then it should not be accepted if a transfer program that results in a smaller loss in efficiency can potentially be used instead. By the same token, if a nontransfer program makes the disadvantaged worse off but results in gains in efficiency, then it should be accepted if there is a transfer program that can potentially compensate the disadvantaged for their losses without fully offsetting the gains in efficiency from the nontransfer program.

This approach requires that internal weights that are derived similarly to the six values that appear in the last column of Table 14.1 be obtained for transfer programs. This has been done by Edward M. Gramlich, who suggests that setting the internal weight for nonrecipients to unity, the internal weight for transfer recipients is on the order of 1.5 to 2. In other words, Gramlich found that it costs taxpayers around $1.50 to $2.00 for each dollar transferred to a recipient under a transfer program.
Gramlich indicates, although these estimates should be considered rough and tentative, if one accepts them as being of the right order of magnitude, then it can be argued that distributitional weights for the disadvantaged should never be set above 1.5 or 2.

Consider, for example, a nontransfer program that costs the advantaged $2.50 for every dollar of benefits received by the disadvantaged. Gramlich’s estimates imply that every dollar received by the disadvantaged under a transfer program would cost the advantaged only $1.50 to $2.00. Thus, in principle, a transfer program could be used instead to make the disadvantaged just as well off as the nontransfer program but at a lower cost to the advantaged. Thus, not only does the nontransfer program have a negative unweighted net social present value, it is inferior to a simple transfer program for redistributing income to the disadvantaged.

Now consider a program that provides the advantaged $2.50 of benefits for every dollar of costs incurred by the disadvantaged. Under these circumstances, each dollar lost under the program by the disadvantaged could, in principle, be reimbursed through a transfer program at a cost to the advantaged of only $1.50 to $2.00. Hence, this program not only has a positive unweighted net social present value, the disadvantaged also can be compensated for their losses without completely offsetting the gains in efficiency from the program.

The argument just presented implies that distributional weights assigned to the disadvantaged should not exceed 1.5 or 2 in value. Higher weights would imply acceptance of inefficient programs that are also inferior to simple transfer programs for redistributing income and rejection of efficient programs that allow the advantaged to enjoy net gains even when the disadvantaged could be fully compensated through income transfers for losses they suffer. Thus, the argument suggests that the two programs in Table 14.1 that have internal weights well in excess of 2 should definitely be accepted, while the two that have internal weights that are close to 2 should probably also be accepted, even though all four programs have adverse effects on the income distribution.

Note, however, that this argument is very similar in spirit to the one underlying the Kaldor-Hicks rule. Both are based on the potential use of transfer payments to compensate losers under a policy, while leaving winners better off than they would be in the absence of the policy. Nothing, however, requires that these transfer payments actually be made.

CONCLUSION

This chapter focuses on the use of distributional weighting to take account of the fact that many policies have divergent impacts on different income groups. Given the absence of generally accepted sets of distributional weights, we suggest that the use of distributional weights should be limited to policies that meet both of the following conditions: (1) they are targeted at the disadvantaged; (2) they result in reductions in overall social efficiency but make low-income persons better off; or they increase social efficiency but make low-income persons worse off.

There may, in fact, be relatively few policies that meet both of these conditions. Those policies that do might be subjected to sensitivity tests based on a plausible
range of weights. Or alternatively, internal weights might be computed, thereby providing policymakers with information on which to base their choice of distributional weights. In either case, however, a cogent argument can be made for not allowing the distributional weights for low-income groups to be set much more than 50 percent to 100 percent above those for higher-income groups.

EXERCISES FOR CHAPTER 14

1. A city is about to build a new sanitation plant. It is considering two sites, one located in a moderately well-to-do neighborhood and the other in a low-income neighborhood. Indeed, most of the residents in the latter neighborhood live below the poverty line. The city's sanitation engineer is adamant that "the city needs the new plant and it has to go somewhere." However, he is indifferent as to which neighborhood it is located in. The plant would operate at the same cost and as efficiently in either neighborhood, and about as many people would be affected by the air pollution emitted by the plant. The city hires an economist to study the two sites. The economist finds that the plant would cause a considerably larger fall in average property values in the well-to-do neighborhood than in the low-income neighborhood, given the more expensive homes that are located in the former. Consistent with this, a contingent valuation study that the economist conducts finds that willingness-to-pay to avoid the sanitation plant is substantially higher in the well-to-do neighborhood than in the low-income neighborhood.

The residents of the poor neighborhood strongly prefer that the plant be built in the well-to-do neighborhood. In the face of the economist's findings, what sort of arguments might they make?

2. Cost-benefit analyses have been conducted of six proposed projects. None of these projects are mutually exclusive and the agency has a sufficient budget to fund those that will make society better off. The findings from the CBAs are summarized here in millions of dollars:

<table>
<thead>
<tr>
<th>Project</th>
<th>Net Social Benefits</th>
<th>Net Group I Benefits</th>
<th>Net Group II Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$2</td>
<td>$2</td>
<td>$0</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>8</td>
<td>-2</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>12</td>
<td>-8</td>
</tr>
<tr>
<td>D</td>
<td>-1</td>
<td>-3</td>
<td>2</td>
</tr>
<tr>
<td>E</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>F</td>
<td>-2</td>
<td>4</td>
<td>-6</td>
</tr>
</tbody>
</table>

Group I consists of households with annual incomes over $15,000, while Group II consists of households with annual incomes under $15,000.

a. According to the net benefit rule, which of these projects should be funded?

b. For which of the projects might distributional considerations be an issue?

c. Compute internal distributional weights for the projects you selected in 2.b. Using these weights, indicate the circumstances under which each project might actually be undertaken.
d. Recompute social net benefits for the six projects using a distributional weight of 1 for Group I and a distributional weight of 2 for Group II. Using these weight-adjusted net social benefit estimates, indicate the circumstances under which each project might actually be undertaken. In doing this, assume that the distributional weight for Group II is an upper bound—that is, it probably overstates society’s true generosity toward low-income households.

NOTES

1 For a further discussion of standing in CBA, see Chapter 2.
3 It would be better to distinguish among the persons or families affected by government programs in terms of their wealth (i.e., the value of their stock of assets), rather than in terms of their income (the observed flow of payments they receive in exchange for the labor, capital, and land that they provide the production process). For example, two households may have similar income, but if one owns a house and the other does not, their standard of living may be quite different. However, income is usually used instead of wealth in categorizing individuals or families because it is more readily measured.
4 An exception sometimes occurs in conducting CBAs in less developed countries, however. Because of the paucity of funds for investment in such countries, some economists argue that each dollar of increase or reduction in savings that results from a project should count more heavily in conducting a CBA of the project than each dollar of increase or reduction in consumption. [See, for example, Anandarup Ray, Cost-Benefit Analysis: Issues and Methodologies (Baltimore: Johns Hopkins University Press, 1984), 15-17.] In principle, however, the relative importance of savings and consumption can be taken into account by using the shadow price of capital approach to discounting, which is described in Chapter 5.
5 This diagram was adapted from Arnold C. Harberger, "On the Use of Distributional Weights in Social Cost-Benefit Analysis," Journal of Political Economy, 86, no. 2 (1978), S87-S120, University of Chicago Press.
8 Arnold Harberger’s classical examination of distributional weighting ("On the Use of Distributional Weights in Social Cost-Benefit Analysis") can be viewed as a critical assessment of whether this assertion provides a basis for conducting distributionally weighted CBA.
9 As discussed by Aidan R. Vining and David L. Weimer, "Welfare Economics as the Foundation for Public Policy Analysis: Incomplete and Flawed but Nevertheless Desirable" The Journal of Socio-Economics, 21, no. 1 (1992), 23-37; John Rawls, A Theory of Justice (Cambridge, MA: Harvard University Press, 1971), and others have provided a more philosophical basis than the reasons listed here for greater income equality.
10 The first two arguments can be derived more formally by specifying a social welfare function. To illustrate, we specify a very simple social welfare function in which individual utility depends upon income, total social welfare depends upon a linear combination of individual utilities, and the possibility of interdependent utility (i.e., one person’s utility being affected by the gains or losses of others) is ignored:

\[ SW = f(u_1(y_1), \ldots, u_n(y_n)) \]

where \( n \) is the total number of individuals in society.
Distributionally Weighted Cost-Benefit Analysis

Totally differentiating the social welfare function yields the following expression:

$$dSW = \sum_{i=1}^{n} \left( \frac{\partial SW}{\partial u_i} \right) (\partial u_i / \partial y_i) dy_i$$

where $dy_i$ represents the change in income resulting from a government policy.

The first argument implies that $\partial u_i / \partial y_i > \partial u_i / \partial y_h$, while the second argument implies that $\partial SW / \partial u_i > \partial SW / \partial u_h$. Thus, together the two arguments imply that $\left( \frac{\partial SW}{\partial u_i} \right) (\partial u_i / \partial y_i) > \left( \frac{\partial SW}{\partial u_h} \right) (\partial u_h / \partial y_i)$.


2From the perspective of social choice theory, this is somewhat naively democratic. See William H. Riker, *Liberalism Against Populism* (San Francisco: Freeman, 1982).

3For an illustration of how this might be done in practice, see D.W. Pearce, *Cost-Benefit Analysis*, p. 71.


5In principle, a set of weights that is consistent with both the diminishing marginal utility of income and the greater income equality arguments could be derived by simply setting them equal to $(\partial SW/\partial u_i)(\partial u_i/\partial y_i)$ for the average member of each group of interest and then dividing the resulting values for each group by the value for the highest-income group. However, finding values for $(\partial SW/\partial u_i)(\partial u_i/\partial y_i)$ is subject to the same informational problems mentioned in the text. Moreover, the term itself was derived in footnote 10 by first arbitrarily specifying a simple social welfare function that did not allow for interdependencies in utility and in which social welfare depends upon a linear combination of individual utilities. As noted in the text, general agreement on the form of the social welfare function does not exist. For example, utilities are probably interdependent and the social welfare function could very well be nonlinear.


8This approach was first developed by Burton A. Weisbrod, "Income Redistribution Effects and Benefit-Cost Analysis" in *Problems in Public Expenditure Analysis*, ed. S.B. Chase (Washington, DC: Brookings Institution, 1968), 177–208.

9The empirical analysis in Weisbrod’s 1968 paper is actually somewhat more complex than suggested in the text. For example, Weisbrod analyzed four water resource projects considered by the U.S. Corps of Engineers. Of these projects, the one with the highest estimated NPV was rejected, while the other three were accepted. Thus, Weisbrod used four equations to estimate weights for four separate sub-groups (rich whites, poor whites, rich nonwhites, and poor nonwhites).

10Table 9.4 also provides estimates of the individual benefit and cost components of the various Work/Welfare Demonstrations.

11This argument was apparently first made by Arnold C. Harberger, "On the Use of Distributional Weights in Social Cost-Benefit Analysis."