6

Cost-Utility Analysis

Stephen Joel Coons
Robert M. Kaplan
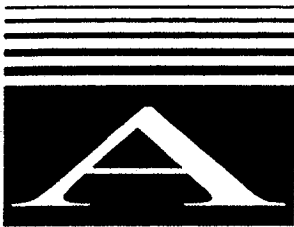
s observed by Neumann and Weinstein,[1] the American public has a love–hate relationship with medical technology. Medical technologies are lauded for saving lives and improving the quality of medical care while, at the same time, they are condemned as the primary cause of the unchecked growth of medical care costs.

Our society's ambivalence stems, in part, from the lack of critical information as to what value is received for the tremendous amount of resources expended on medical care. As Maynard[2] has stated, it is commonplace in health care "for policy to be designed and executed in a data free environment!" Although the implicit objective of medical technology is to improve health outcomes, there is minimal evidence of the true effectiveness of many current healthcare practices.[3] In addition, measures of the overall quality of the US healthcare system, such as access to primary health care, health indicators (e.g., infant mortality, life expectancy), and public satisfaction in relation to costs, provide evidence that we trail other countries that spend significantly less than the US does on medical care.[4]

Earlier chapters have discussed the pressing need to maximize the net health benefit derived from the utilization of limited healthcare resources. Cost-effectiveness analysis (CEA) and cost-benefit analysis (CBA) have been presented as methods for assessing the costs and consequences of healthcare technologies, particularly pharmaceuticals. The purpose of this chapter is to discuss another method for evaluating the value obtained for the money spent: cost-utility analysis (CUA).
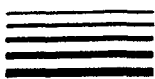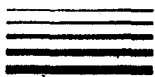
What Is Cost-Utility Analysis?

CUA is a formal economic technique for assessing the efficiency of healthcare interventions. It is considered by some to be a specific type of CEA[5] in which the measure of effectiveness is a utility- or preference-adjusted outcome. However, in this chapter, we will consider it as a separate and distinct economic technique.

CUA is one of the newest, and perhaps most controversial, types of economic evaluation. The controversy stems mainly from the measurement of utility. Utility is the value or worth placed on a level of health status, or improvement in health status, as measured by the preferences of individuals or society.[6] The measurement of utility is necessary for the calculation of the most commonly used outcome measure in this type of analysis: *quality-adjusted life-years (QALYs) gained*. There is no true consensus as to the most appropriate measurement approach. (The measurement of utility is discussed in greater depth later in the chapter.)

Nevertheless, CUA has some distinct advantages over CBA and CEA. CBA suffers from the difficulty of translating all costs and consequences into monetary terms.[7] It is especially difficult to translate patient-reported outcomes (e.g., quality of life) into dollars. In addition, CBA carries the potential for discrimination because it favors treatment for people who are working or those who are more wealthy. CEA is limited by the inability to simultaneously incorporate multiple outcomes from the same intervention or to compare interventions with different outcomes. In CEA, although the outcome measure is in natural units (e.g., life-years saved), no attempt is made to value the consequence or outcome in terms of quality or desirability. In contrast, CUA incorporates the quality of (or preference for) the health outcome achieved.[6] CUA, using QALYs gained as the outcome measure, is the most common approach to combining quantity and quality-of-life outcomes in economic evaluations.[8]

When Is Cost-Utility Analysis Appropriate?

Drummond et al.[6] enumerated several circumstances where CUA may be the most appropriate analytic approach:

1. When quality of life is *the* important outcome. For example, when comparing interventions that are not expected to have an impact on mortality, but a potential impact on patient function and well-being (e.g., treatments for arthritis).

2. When quality of life is *an* important outcome. For example, evaluation of the outcomes associated with the treatment of acute myocardial infarction. Not only is lives saved an important outcome measure, but also the quality of the lives saved (e.g., the impact of a treatment-induced stroke

in a survivor). Another example is the treatment of cancer. A chemotherapeutic agent may increase survival while it decreases the quality of the life being lived.

3. When the intervention affects both morbidity and mortality and a combined unit of outcome is desired. For example, evaluation of a therapy, such as estrogen use by postmenopausal women, that can improve quality of life, may reduce mortality from certain conditions (e.g., heart disease), but may increase mortality from other conditions (e.g., uterine cancer).

4. When the interventions being compared have a wide range of potential outcomes and there is a need to have a common unit of outcome for comparison. This is most commonly the case when a decision-maker must allocate limited resources among interventions that have different objectives and resultant benefits. For example, the choice between providing increased prenatal care or expanding a hypertension screening and treatment program.

5. When the objective is to compare an intervention with others that have already been evaluated in terms of cost per QALY (or equivalent) gained.

The identification, valuation, and measurement of costs is covered elsewhere in this book and will not be repeated here. In addition, Chapter 7 covers quality-of-life assessment. Health-related quality of life is an integral part of CUA; however, not all instruments measuring this component have outcome scores that can be incorporated into CUA. As stated by Hopkins,[9] quality of life measures:

> . . . have proved successful in tracking the effects of medical and surgical interventions and reflecting apparently more realistically the outcomes of these interventions. However, the multidimensional nature of these scales is perceived by some as a disadvantage, as it is difficult to compare outcomes between patients and across procedures. How can there be a 'trade-off,' for example, between reduction in pain and depression of mood? There is therefore considerable interest in attempting to value a health state in terms of a single number. Such valuations can then be integrated with the dimension of time in that state to allow comparisons of values achieved by different interventions in different clinical disorders. The best known of these integrated indices is the quality-adjusted life year . . .

QALYs integrate in a single summary score the net health improvement gains, in terms of both quantity and quality of life, experienced by a group of individuals. Although some economic evaluations reported in the literature have used disease-specific quality-of-life scales or general health profiles as outcome measures, most have incorporated valuations of health state preferences or utilities for the purpose of calculating QALYs.[8] This chapter focuses on assessing the health state utilities needed to calculate QALYs, the CUA ratio's most commonly used denominator.

However, before proceeding, we must clarify some terminology. The term "utilities" used in health state valuation literature does not correspond to the classical use of the same term by economists and philosophers of the 19th century. The current use of the term is derived from von Neumann and

Morgenstern's[10] theory of rational decision-making. Torrance and Feeny[11] suggested that to avoid confusion it would be preferable to call some of the valuations discussed in the following section health state "value preferences" rather than utilities; however, in this chapter, that distinction will not be made. This will be discussed in greater detail below.

Need for Health State Utility Assessment

Health care has different objectives. The objective of the care provided by a diabetologist might be a reduction in diabetic complications. Oncologists strive to keep their patients alive and may be satisfied with a short increase in survival time, whereas primary care providers often focus on shortening the cycle of acute illnesses for which mortality is not an immediate concern. All of these providers are attempting to improve the health of their patients. However, they each measure health in a different way. Comparing the productivity of a diabetologist to that of an oncologist may be like comparing apples to oranges. In other words, there is usually no way to directly compare the productivity of different providers when the intended outcomes are different.

The diversity of objectives and resulting outcomes in health care has led many analysts to focus on the simplest common ground. Typically, that is mortality or life expectancy. When mortality is studied, those who are alive are statistically coded as 1.0 and those who are dead are statistically coded as 0.0. Mortality allows the comparison between different diseases. For example, we can compare the years of life lost from heart disease to the years of life lost from cancer. The difficulty is that everyone who remains alive is given the same score. A person with endstage renal disease is given the same score as someone who is healthy. Utility assessment allows the quantification of levels of wellness on the continuum anchored by death and optimum function.

Conceptual Model

To evaluate health-related quality of life, we must consider all of the different ways that illness and its treatment affect outcomes. It can be said that there are only two central categories of outcomes: life duration and quality of life.[12] We are concerned about any illness or disability because it might make us live a shorter period of time. In addition, we are concerned about the impact of an illness or the effects of its treatment on quality of life. Assessment should consider three basic questions[13]: (1) Does the illness or its treatment shorten life? (2) Does the condition or its treatment make life less desirable and, if so, how much less desirable? (3) What are the duration effects; that is, how much life is lost or how long is the period of diminished quality of life?

Life duration, or quantity of life, as affected by an illness or its treatment is the easier of the concepts to measure. Actuarial mortality data allow for the determination or estimation of the shortened quantity of life. However, the impact of an illness or its treatment on the quality of life is less obvious or objective.

Health-related quality of life is a multidimensional construct. Its general measurement can result in a single outcome score (i.e., health index) or an array of scores for individual quality-of-life dimensions (i.e., health profile). The index and the profile represent the two complimentary approaches to quality-of-life assessment: the decision theory or utility approach and the psychometric approach, respectively.[14] Chapter 7 addresses quality-of-life assessment. This chapter's discussion of quality of life will focus on the incorporation of the utility approach in CUA.

Within the last few years there has been growing interest in using quality-of-life data to help evaluate the cost-effectiveness or cost-utility of healthcare programs. As touched upon earlier, CUA expresses the outcomes of health care in a common outcome unit that is equivalent to a well-year of life. The same outcome has been described as QALYs[15] or healthy years of life.[16] Since the term "QALY" has become most popular, we will use it in this chapter. QALYs integrate mortality, morbidity, and preferences into a comprehensive index number. If a man died of a stroke at age 50 and we would have expected him to live to age 75, it might be concluded that the disease was associated with 25 lost life-years. If 100 men died at age 50 (and also had a life expectancy of 75 years) we might conclude that 2500 life-years (100 men × 25 years) had been lost.

Death is not the only outcome of concern in stroke. For many adults the stroke results in disability over long periods of time. The quality-of-life loss can occur even when life expectancy is unaffected. Quality-of-life consequences of illnesses can be quantified and used to adjust length of life for its quality. For example, a disease that reduces quality of life by half will take away 0.5 QALYs over the course of one year. If it affects two people, it will take away 1.0 QALYs (2 × 0.5) over a one-year period. A medical treatment that improves quality of life by 0.2 for each of five individuals will result in the equivalent of 1.0 QALY if the benefit is maintained over a one-year period. This system has the advantage of considering both benefits and adverse effects of interventions in terms of the common QALY units.

Concept of Relative Importance of Dimensions

Nearly all health-related quality-of-life measures have multiple dimensions. The exact dimensions vary from measure to measure. There is considerable debate about which dimensions need to be included. For example, the most commonly included dimensions are physical functioning, role functioning, and mental health.

Different dimensions might be used to record treatment adverse effects as well as benefits. For example, a medication to control high blood pressure might be associated with low probabilities of dizziness, tiredness, impotence, and shortness of breath. The major challenge is in determining what it means when someone experiences an adverse effect. This requires the effect to be placed within the context of the total health outcome. For example, should a patient with insulin-dependent diabetes mellitus discontinue therapy because of skin irritation at the injection sites? Clearly, local irritation is an adverse effect of treatment. But, without treatment the patient would die. Often the issue is not whether treatment causes adverse effects, but how we should place these effects within the perspective of total health. Ultimately, we must decide whether treatment produces a net benefit or a net deficit in health status.

Many measures of health-related quality of life simply tabulate frequencies for different symptoms or represent health status using profiles of outcomes. Figure 1 is a representation of one such profile.[13] The figure represents three hypothetical treatment profiles. It is common in the presentation of these profiles to connect the points even though increments on the x axis are not meaningful. T-scores (y axis) are standardized scores with a mean of 50 and a standard deviation of 10. Treatment 1 may produce benefits for physical functioning but decrements for role functioning. Treatment 2 may produce decrements for physical functioning but increments for role functioning. This information may be valuable for diagnostic purposes. However, ultimately, clinicians make some general interpretations of the profile by applying a weighting system. They might decide that they are more concerned about physical than role functioning, or vice versa. We must recognize, however, that judgment about the relative importance of dimensions is common. Physicians may ignore a particular test result or a particular symptom because another one is more important to them. Typically, however, it is done arbitrarily. We suggest that the process by which relative importance is evaluated can and should be studied explicitly.

There are a variety of conceptual and technical issues relevant to preference or utility assessment.[17-20] For example, different approaches to preference assessment can yield different results. However, these differences might be expected because the different approaches are based on different underlying conceptual models. As a result, the preference assessment techniques ask different questions. The following sections attempt to elucidate some of these conflicts.
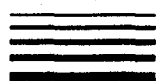
Concept of Utility

The concept of QALYs has been discussed in the literature for nearly 25 years. Perhaps the first application was suggested by Fanshel and
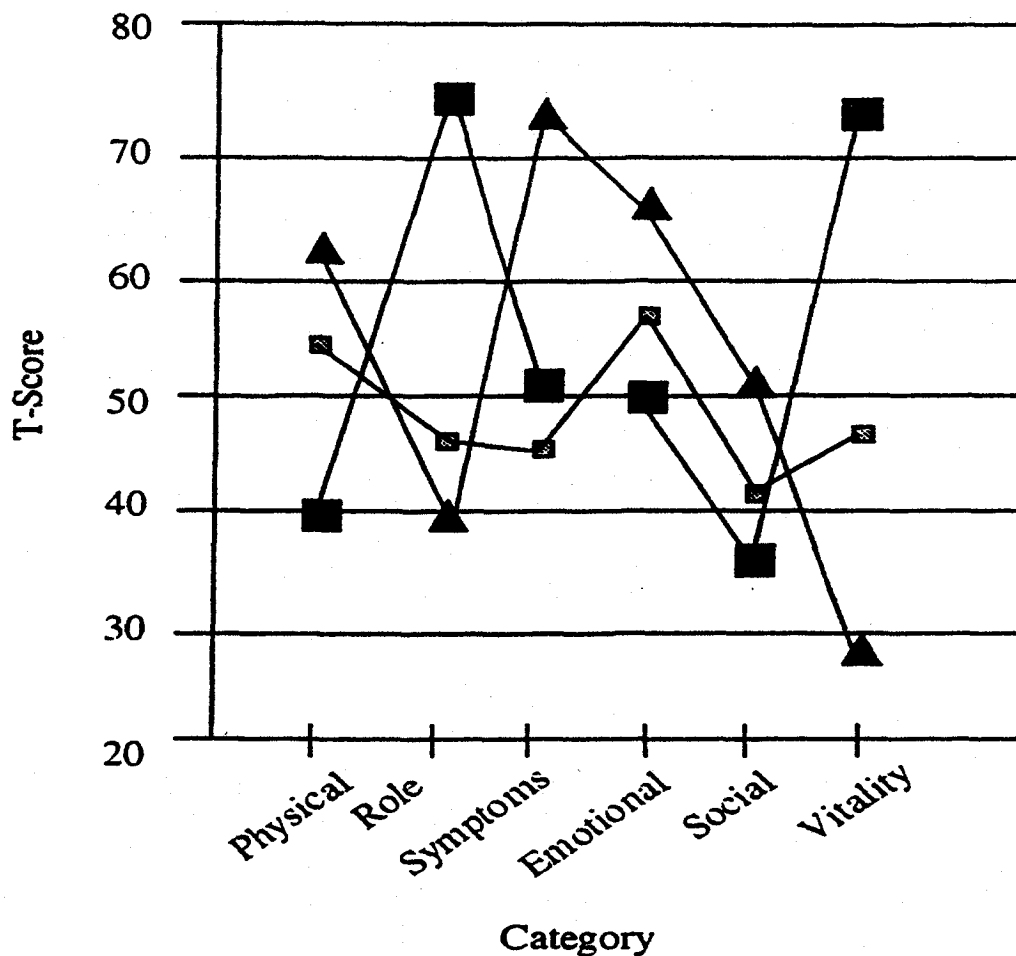
Figure 1. Comparison of profiles from three hypothetical treatments.[13] T-scores are standardized scores with a mean of 50 and a standard deviation of 10. ▲ = treatment 1; ▧ = treatment 2; ■ = treatment 3.

Bush.[21] Soon after, Torrance[22] introduced a conceptually similar model. Since then, a variety of applications have appeared. Although most of these models are conceptually alike, variations between the approaches have led to some inconsistent findings, some of which are highlighted later in the chapter.

Despite the differences in approach, some important assumptions are similar. For example, all of these approaches assume that one full healthy year of life is scored 1.0. Years of life in less than optimal health are scored as less than 1.0. The basic assumption is that two years scored as 0.5 add up to the equivalent of 1.0 year of complete wellness. Similarly, four years scored as 0.25 sum to the equivalent of 1.0 completely well-year of life. A treatment that moves a patient from 0.5 to 0.75 produces the equivalent of 0.25 QALYs. If applied to four individuals, and the duration of the treatment effect is one year, the effect of the treatment would be equivalent to

1.0 completely well-year of life. The disagreement among most researchers is not over the QALY concept but rather over how the weights for cases between 0.0 and 1.0 are obtained. However, that is not to say that there are no concerns about using QALYs in health policy decisions.[23-27] In addition Mehrez and Gafni[28,29] have proposed that the healthy-year equivalent (HYE) is a more appropriate outcome measure than the QALY. They assert that the HYE, like the QALY, combines both quality of life and quantity of life; however, HYEs more fully represent individuals' preferences in the calculation of the trade-offs between quality and quantity of life. A discussion of these concerns is beyond the scope of this chapter.

HISTORY OF THE UTILITY THEORY

The history of the utility theory and its applications to health outcomes assessment has been reviewed by Torrance and Feeny.[11] Health utility assessment has its roots in the work of von Neumann and Morgenstern[10] who published their classic work a half century ago. Their mathematical decision theory characterized how a rational individual should make decisions when faced with uncertain outcomes. They outlined axioms of choice that have been formally evaluated and have become basic foundations of decision analysis in business, government, and health care. Their work has been expanded upon by Raiffa[30] and others.[31,32] Torrance and Feeny[11] emphasized that the use of the term "utility theory" by von Neumann and Morgenstern was unfortunate. Their reference to utility differs from the more common uses by economists that emphasize consumer satisfaction with commodities that are received *with certainty*. Nineteenth century philosophers and economists assumed the existence of cardinal (or interval level) utilities for these functions. A characteristic of cardinal utilities is that they can be aggregated across individuals and used for utilitarian social policy.

By the turn of the century, Pareto challenged the value of cardinal utilities and demonstrated that ordinal utilities could represent consumer choice.[33] In a classic essay, this work was extended by Arrow and Debreu.[34] Arrow[35] had previously argued that there are inconsistencies in individual preferences under certainty and that meaningful cardinal preferences cannot be measured and may not even exist. As a result, most economists maintain that averaged or aggregate preferences have little meaning.

There are several reasons why Arrow's work may not be applicable to the aggregation of utilities in the assessment of QALYs. First, utility expressions for QALYs are expressions of consumer preference *under uncertainty*. The traditional criticisms of microeconomists are directed toward decisions under certainty rather than uncertainty.[11] A second issue is that Arrow assumed that the metric underlying utility was not meaningful and not standardized across individuals. Substantial psychometric evidence now suggests that preferences can be measured using scales that have meaningful interval or ratio properties. When cardinal (interval) util-

ities are used instead of rankings, many of the potential problems in the Impossibility Theorem are avoided.[36,37]

It is also important to recognize that different approaches to the calculations of QALYs are based on very different underlying assumptions. One approach considers the duration someone is in a particular health state as conceptually independent from the utility for the state.[12,15] The other approach merges duration and utility.[11] This distinction is central to the understanding of the difference in approaches and the required evidence for the validity of the utility assessment procedure.

In the approach advocated by Kaplan and Anderson[12] and Weinstein and Stason,[15] utilities for health states are obtained at a single point in time. For example, persons in a particular health state, such as confinement to wheelchair, who performed no major social role are asked to assess the utility of that health state. Suppose that this state is assigned a value of 0.5. Then, patients in this state are observed over the course of time to empirically determine their transitions to other states of wellness. If they remain in the state for one year, then they would lose the equivalent of 0.5 well-years of life. The key to this approach is that the preferences only concern a single point in time and that the transition is determined through observation or expert judgment. The alternative approach emphasized by Torrance and Feeny[11] and Nord[38] obtains preference for both health state and for duration. These approaches also consider the more complex problems of uncertainty. Thus, they are consistent with the von Neumann and Morgenstern's[10] notion of decision under uncertainty in which probabilities and trade-offs are considered explicitly by the judge.

Methods for Assessing Utility

CUA requires an assessment of utilities for health states. A variety of different techniques have been used to assess these utilities. These techniques will be summarized briefly. Then, comparisons between the techniques will be considered. Some analysts do not measure utilities directly. Instead, they evaluate health outcome by simply assigning a reasonable utility.[39] However, most current approaches have respondents assign weights to different health states on a scale ranging from 0 (for dead) to 1.0 (for wellness). The most common techniques include category rating scales, magnitude estimations, the standard gamble, the time trade-off, and the equivalence person trade-off. Each of these methods will be described briefly.

RATING SCALES

Rating scales require the respondent to assign a numeric value to objects. There are several methods for obtaining rating scale information. The category scale, exemplified by the familiar 10-point rating scale, is

efficient, easy to use, and applicable in a large number of settings. Typically, the subjects read the description of a case and rate it on a 10-point scale ranging from 0 for dead to 10 for asymptomatic optimum function. The endpoints of the scale are typically well defined.

Another common rating scale method is the visual analog scale. The visual analog method shows a subject a line, typically 100 centimeters in length, with the endpoints well defined. The subject's task is to mark the line to indicate where their preference rests for one or more health states in relation to the two poles.

Appropriate applications of rating scales reflect contemporary developments in cognitive sciences. Judgment-decision theory has been dominated by the belief that human decisions follow principles of optimality and rationality. A considerable amount of research has challenged the normative models that have attempted to demonstrate rational choice. The development of cognitive theories, such as information integration theory,[40] provide better explanations of the cognitive process of judgment. Information integration theory includes two constructs: integration and valuation. A large body of evidence indicates that rating scales provide meaningful metrics for the expression of these subjective preferences.[40] Although there have been some challenges to the use of rating scales, most biases can be overcome with the use of just a few simple precautions, such as clear definitions of the endpoints and preliminary practice with cases that make the endpoints salient.[40]

MAGNITUDE ESTIMATION

Magnitude estimation is a common psychometric method that is believed by psychophysicists to yield ratio scale scores. In magnitude estimation, a specific case is selected as a standard and assigned a particular number. Then, other cases are rated in relation to the standard. Suppose, for example, the standard is assigned the number 10. If a case is regarded as half as desirable as the standard, it is given the number 5. If it is regarded as twice as desirable, it is given the number 20. Ratings across subjects are standardized to a common metric and aggregated using the geometric mean. Advocates for magnitude estimation argue that the method is meaningful because it provides a direct estimate of the subjective ratio. Thus, they believe, the magnitude estimate has the properties of a ratio scale. However, magnitude estimation has been challenged on several grounds. The method is not based on any specific theory of measurement and gains credibility only through face validity.[40] Further, the meaning of the scores has been challenged. For example, the values are not linked directly to any decision process. What does it mean if one case is rated as half as desirable as another? Does it mean that the respondent would be indifferent between a 50–50 chance of the higher valued outcome and a certainty of the alter-
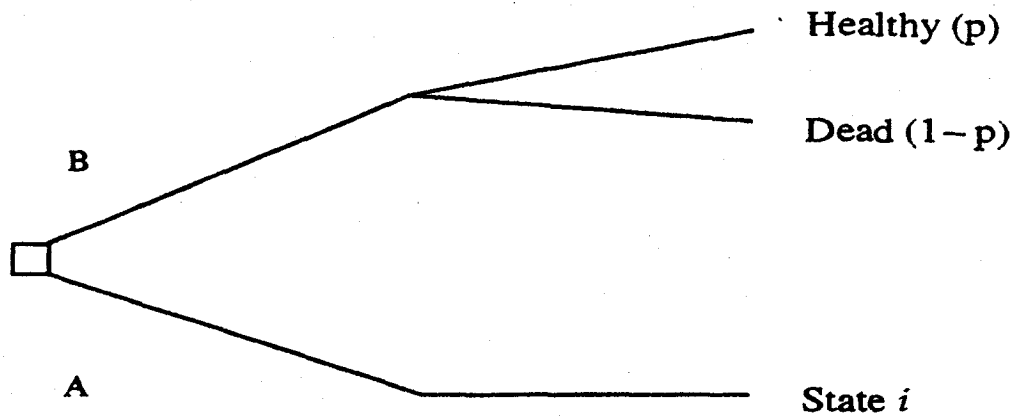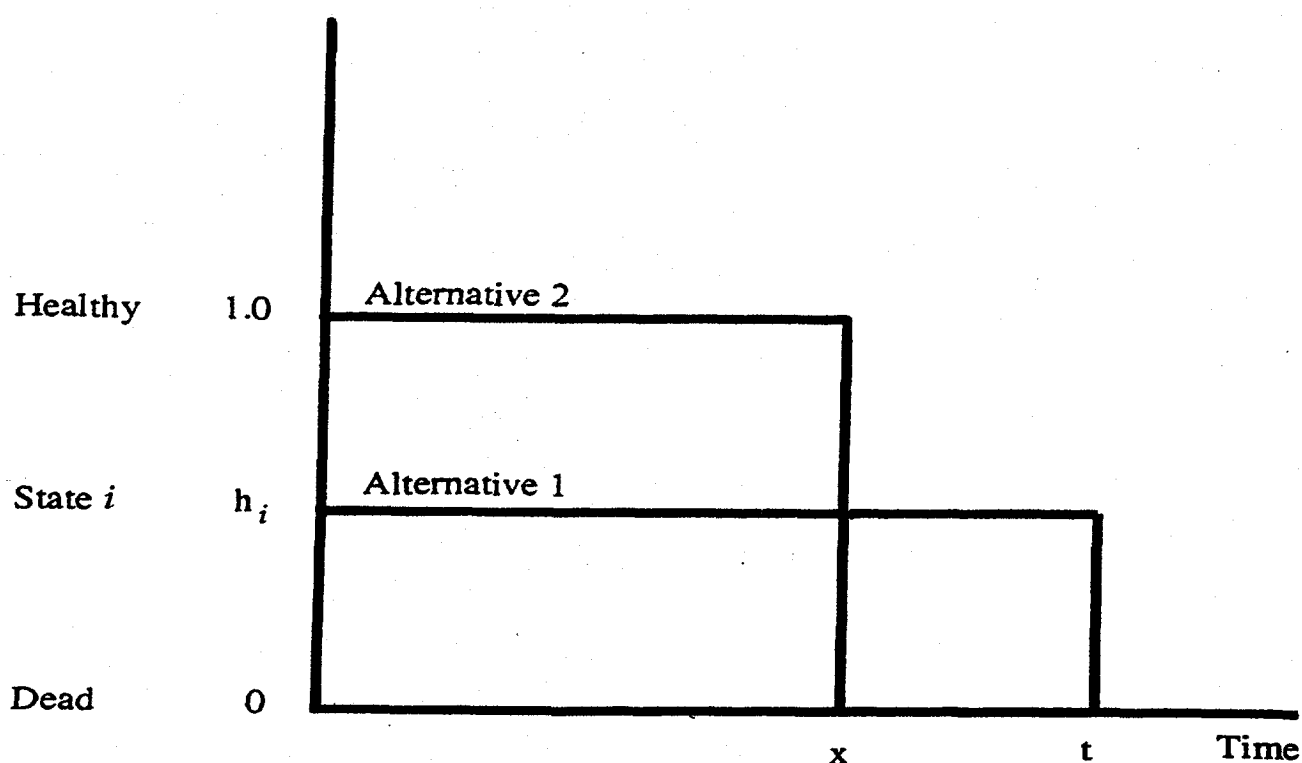
Healthy (p)

Dead (1 − p)

B

A

State *i*

Figure 2. Standard gamble for a chronic health state. *i* = chronic health state; p = probability of achieving perfect health.

native valued as half as desirable? These issues have not been systematically addressed in the health status literature.

STANDARD GAMBLE

Category rating and magnitude estimation are methods commonly used by psychometricians. Typically, the tasks emphasize wellness at a particular point in time and do not ask subjects to make trades or to consider aspects of uncertainty. Several methods more explicitly consider decisions under uncertainty. The standard gamble offers a choice between two alternatives: choice A—living in health state *i* (a chronic health state between perfect health and death) with certainty, or choice B—taking a gamble on a new treatment for which the outcome is uncertain. Figure 2 shows this trade. The respondent is told that a hypothetical treatment will lead to perfect health with a probability of p or immediate death with a probability of 1 − p. They can choose between remaining in state *i* that is intermediate between wellness and death or taking the gamble and trying the new treatment. The probability is varied until the subject is indifferent between choices A and B. For example, if a subject is indifferent between choices A and B when p = 0.65, the utility of state *i* is 0.65.

The standard gamble has been attractive because it is based on the axioms of utility theory. The choice between a certain outcome and a gamble conforms to the exercises originally proposed by von Neumann and Morgenstern.[10] Although the interval properties of the data obtained using the gamble have been assumed, they have not been empirically demonstrated.[17] A variety of other problems with the gamble also have become apparent. For example, it has often been stated that the standard gamble has face validity because it approximates choices made by medical patients.[41] However, treatment of most chronic diseases does not approximate the

Figure 3. Time trade-off for a chronic health state. $h_i = x \div t$, where h_i = preference value or utility for state i; state i = chronic health state; t = life expectancy for an individual with chronic health state i; and x = time at which respondent is indifferent between alternatives 1 and 2.

gamble. There is no known product that will cure a patient with arthritis nor is there one that is likely to kill her. In other words, the decision-making experience of the patient is not likely to include an option that has a realistic gamble. Further, the cognitive demands of the task are high.

TIME TRADE-OFF

The concept of probability is difficult for most respondents and requires the use of visual aids or props to assist in the interview. Thus, an alternative to the standard gamble, which is also consistent with the von Neumann and Morgenstern[10] axioms of choice, uses a trade-off in time.[42] Figure 3 demonstrates the trade-off for a chronic disease state. Here, the subject is offered a choice of living in health state i (chronic health state considered better than death but less desirable than perfect health) for time t (life expectancy for an individual with chronic health state i) or perfect health for time x. Time x and t are followed by immediate death. Time x is varied until the respondent is indifferent between the two alternatives. Presumably, all subjects would choose a year of wellness versus a year with

some health problem. However, by reducing the time of wellness and leaving the time in the suboptimal health state fixed (such as one year), an indifference point can be determined ($h_i = x \div t$) (h_i is the utility or preference value for chronic health state i). For example, a subject may rate being in a wheelchair for two years as equivalent to perfect wellness for one year ($1 \div 2 = 0.5$). The time trade-off is theoretically appealing because it is conceptually equivalent to a QALY.

PERSON TRADE-OFF

Finally, a person trade-off technique allows comparisons of the numbers of people helped in different states. For example, respondents might be asked to evaluate the equivalencies between the number of persons helped by different programs. They might be asked how many persons in state B must be helped to provide a benefit equivalent to helping one person in state A. From a policy perspective, the person trade-off also directly seeks information similar to that required as the basis for policy decisions.

Differences Between the Methods

Several papers in the literature have compared utilities for health states as captured by different methods. These differences have been reviewed by Nord.[43] In general, standard gamble and time trade-off methods give higher values than rating scales in most, but not all studies. In about half of the studies reported, time trade-off yields lower utilities than standard gamble. In one of the earlier studies, Patrick et al.[44] found that person trade-off methods gave the same results as rating scales. However, these findings were not confirmed in more recent studies.[38] Magnitude estimation has produced results that are highly inconsistent across studies.[43]

The variability of health state utilities in comparisons of different studies is hardly surprising. The methods differ substantially in the questions posed to respondents. In summary, there is substantial debate about which technique should be used to acquire utility information. Results obtained from different methods do not correspond, although they typically have a high degree of similarity in the ranks they assign to outcomes. However, the differences in preferences yielded by different methods can result in different allocations of resources if the preferences are not obtained on a linear or interval response scale. For example, suppose that the difference between the effect of a drug and a placebo is 0.05 units of well-being as assessed by rating scales and 0.02 as measured by magnitude estimation. The benefit would have to last 20 years to produce 1.0 QALY if rating scale utilities were used, and 50 years if magnitude estimation utilities were used. Aggregation of benefits necessarily requires an underlying linear response scale in which equal differences at different points along the response scale

are equally meaningful. For example, the differences between 0.2 and 0.3 (0.1 QALY if the duration is one year) must have the same meaning as the difference between 0.7 and 0.8. A treatment that improves a patient's condition from 0.2 to 0.3 must be considered of equal benefit to a treatment that improves it from 0.7 to 0.8. Confirmation of this scale property has been demonstrated for rating scales but not for other methods.[40,45]

Another difference between methods is the inclusion of information about uncertainty in the judgment process. Time trade-off, standard gamble, or person trade-off all theoretically include some judgment about duration of stay in health state. Magnitude estimation and rating scales typically separate utility at a point in time from probability. Considerably more theoretical and empirical work will be necessary to resolve these differences of approach.

Nord[46] recently argued for quality assurance standards for QALY calculations. These recommendations were based on a review that revealed inconsistency in the methods used to assess utilities. According to Nord, the utility assessment typically lacked a theoretical or empiric basis. Apparently, this refers to inattention to economic theories. However, others have noted that some utility assessment approaches are based on different theoretical models and empiric results. For example, advocates for the use of rating scales offer evidence that the methods found theoretically and empirically justified by economists fail to meet the basic requirements for an interval response scale.[47] Nord et al.[48] suggested that the person trade-off be used as the standard against which other methods are compared. However, they did not offer evidence that data obtained using the person trade-off meet standards of reliability, validity, and interval scale property. The person trade-off does not meet the face validity criterion of being a direct estimate of a QALY. However, it is not clear that subjects do not make other cognitive errors when applying the method. This is likely to remain an area of active debate into the near future.

Whose Utilities or Preferences Should be Used?

Choices between alternatives in health care necessarily involve preference judgment. For example, the inclusion of some services in a basic benefits package and the exclusion of others is an exercise in value, choice, or preference. There are many levels at which preference is expressed in the healthcare decision process. For example, an older woman may decide to cope with the symptoms of upset stomach in order to gain relief from the discomfort of osteoarthritis. A physician may order pelvic ultrasound to ensure against missing the very low probability that a 40-year-old woman has ovarian cancer. Or, an administrator may decide to put resources into prevention for large numbers of people instead of devoting the same resources to organ transplants for a smaller number.

In CUA, preferences are used to express the relative importance of various health outcomes. There is a subjective or qualitative component to health outcome. Whether we prefer a headache or an upset stomach caused by its treatment is a value judgment. Not all symptoms are of equal importance. Most patients would prefer mild fatigue (an adverse effect of treatment) to a severe headache (the symptom eradicated by treatment). Yet, providing a model of how well treatments work implicitly includes these judgments. Models require a precise numeric expression of this preference. CUA explicitly includes a preference component to represent these trade-offs.

Some models obtain preferences from random samples of the general population.[37] It is recognized that administrators ultimately choose between alternative programs. Community preferences may represent the will of the general public and not those of administrators. Yet there is considerable debate about technical aspects of preference assessment. Some of the debate has to do with whose preferences are considered.

In most areas of preference assessment, it is easy to identify differences between different groups or different individuals. It might be argued that judgments about net health benefits for white men should not be applied to Hispanic men who may give different weight to some symptoms. Preferences for movies, clothing, or political candidates differ for social and cultural groups; it is assumed that these same differences extend to health states. Allocation of resources to Medicaid recipients, for example, would be considered inappropriate when the preferences came from both Medicaid recipients and nonrecipients.[49] Other analysts have suggested that preference weights from the general population cannot be applied to any particular patient group. Rather, patient preferences from every individual group must be obtained.

Most studies do not support the common belief that preferences differ. Some small, but significant differences between demographic groups have been observed.[50] Studies have found little evidence for preference difference between patients and the general population. For example, Balaban et al.[51] compared preference weights obtained from patients with arthritis with those obtained from the general population in San Diego. They found a high degree of correspondence for ratings of cases involving patients with arthritis. Similar results were found by Hughes et al.[52] among HIV-infected patients. Studies of patients with cancer have had comparable findings (unpublished study: Nerenz DR, Golob K, Trump DL. Preference weights for the Quality of Well-Being Scale as obtained from oncology patients. Henry Ford Hospital, Detroit, MI, 1990). Studies also have shown a high degree of similarity in preferences provided by men versus women, the medically insured versus the uninsured, those ever in wheelchairs versus those never in wheelchairs, British versus Americans, citizens of Oregon versus those of California, and residents of three different European communities.[53]

It would be incorrect to say that there are never any mean differences in preference, since significant differences in preferences have been observed in several studies. However, these differences were typically small.[53] Further analysis will be required to determine whether these differences affect the conclusions of various analyses.

A related problem is the assumption that all people in the same health state should get the same score. Most approaches to utility assessment use the mean preference for a particular case to represent all individuals who meet a common definition. For example, suppose that the average utility for being in a wheelchair, limited in major activities, and having missing limbs is 0.50. The models would assign the same number to all individuals who occupy that state. However, there is substantial variability in how individuals view their own health. If individual preferences are used, there might be significant variation in scores across people with identical objective descriptions.[54] Despite the appeal of individualized preferences, they rarely lead to different treatment decisions than would be obtained from the use of aggregate preferences.[55]

Multiattribute Health Status Classification Systems

Although it is important to understand the various approaches to the measurement of health state utilities/preference values, some pharmacoeconomic researchers conducting a CUA will not measure health state utilities directly. They may use one of the existing multiattribute health status classification systems for which the utility functions have been empirically derived. Two such instruments developed in North America are the Quality of Well-Being Scale (QWB) developed at the University of California–San Diego, La Jolla, CA, and the Health Utilities Index (HUI) developed at McMaster University, Hamilton, Ontario, Canada.

The QWB is a general quality-of-life instrument that includes symptoms or problems plus three dimensions of functional health status: mobility, physical activity, and social activity. Standardized preference weights for the QWB have been measured (using the category rating scale method) and validated on a general population in San Diego.[56] Other investigators have reweighted the symptoms/problems and function levels of the QWB in specific populations, such as patients with arthritis[51] and HIV-infected subjects,[52] and have found the generalizability of the original weights to be very high.

The HUI is another general instrument that describes the health status of a person at a point in time in terms of ability to function on a set of attributes or dimensions of health status. The original version (Mark I) of the HUI consists of four attributes and a formula to calculate utilities.[57] The second version (Mark II) consists of seven attributes and formulas for the calculation of utilities and preference values.[58] The measurement of

the preferences/utilities for the health status classification system were made with visual analog scales and the standard gamble technique. The most recent and potentially most useful version, Mark III, has eight attributes: vision, hearing, speech, ambulation, dexterity, cognition, pain and discomfort, and emotion.[59] The utility/preference functions are not yet available, but those for the Mark II can be used in the interim.[a]

Two additional potentially useful multiattribute health status systems in the latter stages of development in Europe are the EuroQol[60] and the Index of Health-Related Quality of Life.[61] The EuroQol was developed concurrently in five languages (Dutch, English, Finnish, Norwegian, and Swedish) by a multidisciplinary team of European researchers.[60] (Spanish and Catalan translations are now available and French, German, and Italian versions are in preparation.) The EuroQol instrument was designed to be self-administered and short enough to be used in conjunction with other measures. It has two parts: a visual analog scale on which patients rate their own health on a scale of 0 to 100, and a questionnaire that classifies subjects into one of 243 health states. The current EuroQol health status classification system has five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has three levels. An earlier version of the instrument had six dimensions and classified individuals into three levels of mobility, self-care, and pain, and into two levels of main activity, family/leisure activity, and anxiety/depression.[62] The EuroQol is truly an instrument in development and a great deal of research is ongoing, and necessary, to demonstrate its use and usefulness.[63-65]

The Index of Health-Related Quality of Life is a measure of social, psychological, and physical functioning that is in development in the UK.[61] It is based on a five-level multidimensional classification system and provides a health profile as well as a unidimensional health index value (0 to 1). The health index value or global score is comprised of three primary dimensions: disability, physical discomfort, and emotional distress. The dimensions are further subdivided into attributes, then scales, then descriptors. The complex and multilevel valuation approach, which incorporated both standard gamble and category rating techniques, is addressed elsewhere.[66] The validity, reliability, and usefulness of the instrument is yet to be determined.

Cost-Utility Analysis and Healthcare Interventions

Figure 4 presents a graphic representation of a hypothetical case in which treatment has increased both the quantity and quality of life for a
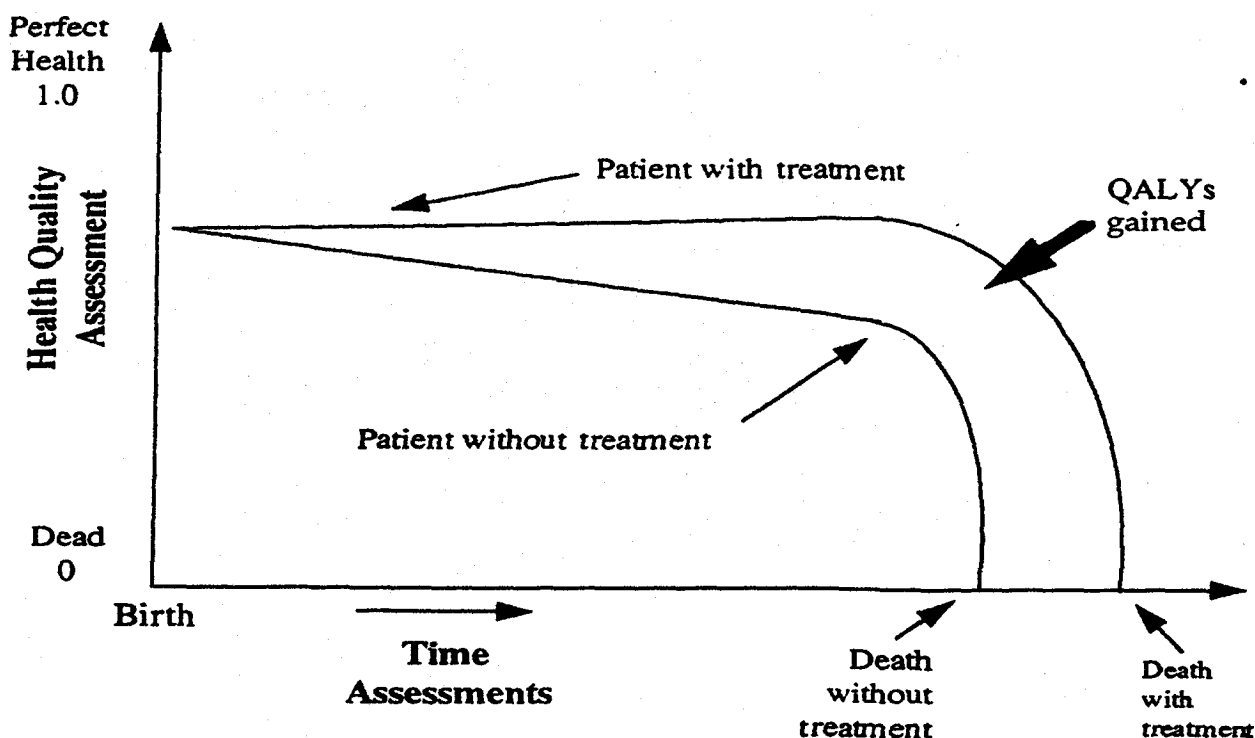
Figure 4. Representation of quality-adjusted life-years (QALYs) gained through a hypothetical treatment.

patient or group of patients. The area between the curves would be calculated to measure the QALYs gained. To complete the CUA, the incremental cost of providing the treatment would be measured and divided by the QALYs gained. Table 1, adapted from Detsky and Naglie,[67] illustrates the different elements that are needed to conduct CEA and CUA when comparing two alternative interventions. An assumption in the table is that quality of life (utility) remains constant over the full life expectancy.

Table 1. Economic Analysis of Two Alternative Treatment Interventions[67]

INTERVENTION	COST ($)	EFFECTIVENESS (LIFE EXPECTANCY) (y)	HEALTH STATE (UTILITY)	QALYs
Treatment A	20,000	4.5	0.60	2.7
Treatment B	10,000	3.5	0.72	2.5

$$\text{Incremental cost-effectiveness ratio} = \frac{\$20,000 - \$10,000}{4.5\,y - 3.5\,y} = \$10,000 \text{ per life-year gained}$$

$$\text{Incremental cost-utility ratio} = \frac{\$20,000 - \$10,000}{2.7\,\text{QALYs} - 2.5\,\text{QALYs}} = \$50,000 \text{ per QALY gained}$$

QALY = quality-adjusted life-year.

Table 2. Quality-Adjusted Life-Years Gained:
Some Tentative Estimates[2]

INTERVENTION	COST/QALY (£)[a]
Cholesterol testing and diet therapy only (all adults, aged 40-69 y)	220
Neurosurgical interventions for head surgery	240
General practitioner advice to stop smoking	270
Neurosurgical intervention for subarachnoid hemorrhage	490
Antihypertensive therapy to prevent stroke (ages 45-64 y)	940
Pacemaker implantation	1,100
Valve replacement for aortic stenosis	1,140
Hip replacement	1,180
Cholesterol testing and treatment	1,480
Coronary artery bypass graft (left main vessel disease, severe angina)	2,090
Kidney transplant	4,710
Breast cancer screening	5,780
Heart transplantation	7,840
Cholesterol testing and treatment (incrementally) of all adults 25-39 y	14,150
Home hemodialysis	17,260
Coronary artery bypass graft (one vessel disease, moderate angina)	18,830
Continuous ambulatory peritoneal dialysis	19,870
Hospital hemodialysis	21,970
Epoetin alfa therapy for anemia in patients undergoing dialysis (assuming a 10% reduction in mortality)	54,380
Neurosurgical intervention for malignant intracranial tumors	107,780
Epoetin alfa therapy for anemia in patients undergoing dialysis (assuming no reduction in mortality)	126,290

QALY = quality-adjusted life-year.
[a]British pounds as of August 1990.

A common method of summarizing and comparing the results of CUAs is in the form of a league table. Table 2 shows a league table compiled by Maynard.[2] There are many concerns regarding the manner in which league tables are constructed and applied.[68,69] As observed by Mason et al.[68] the source studies in league tables often use various means of calculating QALYs and different years of origin, discount rates, settings, and types of comparison programs. For example, a study by Hornberger et al.[70] compared six methods for deriving cost/QALY data for patients receiving in-center hemodialysis. Results from the 58 patients in their study demonstrated a range of $34,893 to $45,254 per QALY gained based on the Sickness Impact Profile[71] and standard gamble technique, respectively. Nevertheless, Mason et al.[68] concluded that although league tables have serious limitations, the systematic comparisons that league tables provide are preferable to the alternative: the reliance on the informal, unsystematic assessments made in the absence of data.

IMPACT OF PHARMACEUTICAL INTERVENTIONS

For pharmaceuticals, as with most other healthcare interventions, the ultimate therapeutic endpoint or outcome is the enhancement of quality of life and/or length of life. Therefore, in theory, the most appropriate outcome measure would be QALYs. There are a number of published studies that have used QALYs and the cost-utility approach to evaluate the economic efficiency of healthcare programs/interventions.[72-74] However, there are very few published reports of the impact of pharmaceutical interventions in which QALYs were an outcome measure. In addition, those studies that have been published are limited by their methodologic approach to the measurement of QALYs. In some cases, QALYs are estimated/modeled from cross-sectional data rather than measured prospectively.

A number of examples of the use of CUA in the evaluation of cancer chemotherapy have appeared in the literature.[75-77] The study by Smith et al.[77] illustrates the importance of adjusting length of life/survival for quality. The authors compared the incremental costs per QALY for surgery plus adjuvant chemotherapy versus surgery alone in patients with colon cancer. They estimated that 2.4 unadjusted life-years were gained from the addition of chemotherapy; however, after adjusting for quality of life, only 0.4 QALYs were gained. The costs of surgery alone and surgery plus chemotherapy were $6000 and $13,000 per patient per 12 months of treatment, respectively. Since the incremental cost of adding the chemotherapy was $7000, the calculated cost per life-year gained was $2916 and the cost per QALY gained was $17,500. The findings are limited by the study's small sample size and other methodologic weaknesses (e.g., measurement of QALYs). However, it demonstrates the potential power of appropriately conducted CUAs in evaluating the outcomes of pharmaceutical interventions more comprehensively than through other approaches (e.g., CEA).

Summary

In a society in which healthcare resources are limited, it is essential that the resources available are used efficiently and equitably. However, for this system to function effectively, data about costs and outcomes are essential. Resources should be used for programs that produce the greatest benefit for the greatest number of people. The lack of good information about input—output relationships in health care has lead to enormous variations in costs and clinical practice patterns.[5] There has been little consensus on what constitutes good clinical practice. The integration of data on the quality of life with corresponding data on life expectancy yields a single index of health benefit, expressed in terms of QALYs. Our interest in life expectancy and quality of life arises from the fact that health care can influence either or both of these.

It is likely that the use of CUA will increase as the need to evaluate the benefits derived from very different healthcare interventions increases. This type of analysis will assist in ensuring that resources are allocated as efficiently as possible to serve health outcome goals. Resources will never be sufficient to provide all the health care that might be given; there are finite resources and potentially infinite demands. CUA provides a systematic approach to comparing ways of using the resources most efficiently in the process of meeting those demands.

Pharmaceutical interventions are a critical component of health care. Pharmaceuticals can produce QALYs by lengthening life, improving the quality of life, or both. As seen in the cancer treatment example, chemotherapy can be a double-edged sword: it can lengthen life while decreasing its quality. However, it is more common for pharmaceuticals to be prescribed to improve the quality of life in people who have non–life-threatening but potentially debilitating conditions (e.g., arthritis, glaucoma). If conducted properly, CUA can be a powerful tool to more comprehensively evaluate the overall impact of pharmacotherapy. This type of research is needed to document the relative value of pharmaceuticals when compared with other medical interventions and to inform decisions as to the most efficient use of finite healthcare resources.

References

1. Neumann PJ, Weinstein MC. The diffusion of new technology: costs and benefits to health care. In: Gelijns AC, Halm EA, eds. The changing economics of medical technology. Washington, DC: National Academy Press, 1991:21-34.

2. Maynard A. Developing the health care market. Econ J 1991;101(408):1277-86.

3. Roper WL, Winkenwerder W, Hackbarth GM, Krakauer H. Effectiveness in health care: an initiative to evaluate and improve medical practice. N Engl J Med 1988;319: 1197-202.

4. Starfield B. Primary care and health: a cross-national comparison. JAMA 1991;266: 2268-71.

5. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practice. N Engl J Med 1977;296:716-21.

6. Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. New York: Oxford University Press, 1987.

7. Patrick DL, Erickson P. Health status and health policy: allocating resources to health care. New York: Oxford University Press, 1993.

8. Drummond M. The role and importance of quality of life measurements in economic evaluations. Br J Med Econ 1992;4:9-16.

9. Hopkins A. Editor's introduction. In: Measures of the quality of life and the uses to which such measures may be put. London: Royal College of Physicians of London, 1992:iii-iv.

10. von Neumann J, Morgenstern O. Theory of games and economic behavior. Princeton, NJ: Princeton University Press, 1944.

11. Torrance GW, Feeny D. Utilities in quality-adjusted life years. Int J Technol Assess Health Care 1989;5:559-75.

12. Kaplan RM, Anderson JP. The general health policy model: an integrated approach. In: Spilker B, ed. Quality of life assessments in clinical trials. New York: Raven Press, 1990:131-49.

13. Kaplan RM, Coons SJ. Relative importance of dimensions in the assessment of health-related quality of life for patients with hypertension. Prog Cardiovasc Nurs 1992; 7:29-36.

14. Coons SJ, Kaplan RM. Assessing health-related quality of life: application to drug therapy. Clin Ther 1992;14:850-8.

15. Weinstein MC, Stason WB. Hypertension: a policy perspective. Cambridge, MA: Harvard University Press, 1976.

16. Russell LB. Is prevention better than cure? Washington, DC: The Brookings Institution, 1986.

17. Froberg DG, Kane RL. Methodology for measuring health state preferences, II: scaling methods. J Clin Epidemiol 1989;42:459-71.

18. Froberg DG, Kane RL. Methodology for measuring health state preferences, III: population and context effects. J Clin Epidemiol 1989;42:585-92.

19. Froberg DG, Kane RL. Methodology for measuring health state preferences, IV: progress and a research agenda. J Clin Epidemiol 1989;42:675-85.

20. Froberg DG, Kane RL. Methodology for measuring health state preferences, I: measurement strategies. J Clin Epidemiol 1989;42:345-52.

21. Fanshel S, Bush JW. A health-status index and its applications to health-services outcomes. Operations Res 1970;18:1021-66.

22. Torrance GW. Social preferences for health states. An empirical evaluation of three measurement techniques. SocioEcon Plan Sci 1976;10:129-36.

23. Carr-Hill RA. Allocating resources to health care: is the QALY (quality adjusted life year) a technical solution to a political problem? Int J Health Serv 1991;21:351-63.

24. Coast J. Developing the QALY concept: exploring the problems of data acquisition. PharmacoEconomics 1993;4:240-6.

25. Harris J. Life: quality, value and justice. Health Policy 1988;10:259-66.

26. LaPuma J, Lawlor EF. Quality-adjusted life-years: ethical implications for physicians and policymakers. JAMA 1990;263:2917-21.

27. Loomes G, McKenzie L. The use of QALYs in health care decision making. Soc Sci Med 1989;28:299-308.

28. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. Med Decis Making 1989;9:142-9.

29. Mehrez A, Gafni A. The healthy-years equivalents: how to measure them using the standard gamble approach. Med Decis Making 1991;11:140-6.

30. Raiffa H. Decision analysis: introductory lectures on choices under uncertainty. Reading, MA: Addison-Wesley, 1968.

31. Bell DE, Farquhar PH. Perspectives on utility theory. Operations Res 1986;34:179-83.

32. Howard RA. Decision analysis: practice and promise. Manag Sci 1988;34:679-95.

33. Parsons T. The structure of social action. Vol. 1. New York, Free Press, 1937:241-9.

34. Arrow KJ, Debreu G. Existence of equilibrium for a competitive economy. Econometrica 1954;22:265-90.

35. Arrow KJ. Social choice and individual values. New York: Wiley, 1951.

36. Keeney RL. A group preference axiomatization with cardinal utility. Manag Sci 1976;23:140-5.

37. Kaplan RM, Anderson JP, Ganiats TG. The Quality of Well-Being Scale: rationale for a single quality of life index. In: Walker SR, Rosser RM, eds. Quality of life assessment: key issues in the 1990s. Lancaster, UK: Kluwer Academic Publishers, 1993:65-94.

38. Nord E. The validity of a visual analogue scale in determining social utility weights for health states. Int J Health Plan Manag 1991;6:234-42.

39. Weinstein MC, Stason WB. Cost-effectiveness of coronary artery bypass surgery. Circulation 1982;66(suppl 3):56-66.

40. Anderson NH. Contributions to information integration theory. Vol. 1–3. Hillsdale, NJ: Erlbaum Publishers, 1991.

41. Mulley AJ. Assessing patient's utilities: can the ends justify the means? Med Care 1989;27(suppl):S269-81.

42. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programmes. Health Serv Res 1972;7:118-33.

43. Nord E. Methods for quality adjustment of life years. Soc Sci Med 1992;34:559-69.

44. Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. Health Serv Res 1973;8:228-45.

45. Hughes TE, Coons SJ, Kaplan RM, Draugalis JR. Evaluation of linearity of the Quality of Well-Being Scale using functional measurement theory (abstract). Qual Life Res 1994;3:47-8.

46. Nord E. Toward quality assurance in QALY calculations. Int J Technol Assess Health Care 1993;9:37-45.

47. Kaplan RM, Feeny D, Revicki DA. Methods for assessing relative importance in preference based outcome measures. Qual Life Res 1993;2:467-75.

48. Nord E, Richardson J, Macarounas-Kirchmann K. Social evaluation of health care versus personal evaluation of health states. Int J Technol Assess Health Care 1993;9:463-78.

49. Daniels N. Is the Oregon rationing plan fair? JAMA 1991;265:2232-5.

50. Kaplan RM, Bush JW, Berry CW. The reliability, stability, and generalizability of a health status index. In: Proceedings of the American Statistical Association. Washington, DC: ASA Social Statistics Section, 1978:704-9.

51. Balaban DJ, Fagi PC, Goldfarb NI, Nettler S. Weights for scoring the quality of well-being instrument among rheumatoid arthritics. Med Care 1986;24:973-80.

52. Hughes TE, Coons SJ, Kaplan RM, Draugalis JR. Reweighting the Quality of Well-being Scale in HIV-infected subjects (abstract). Qual Life Res 1994;3:79-80.

53. Kaplan RM. The hippocratic predicament: affordability, access, and accountability in health care. San Diego: Academic Press, 1993.

54. O'Connor GT, Nease RF Jr. The descriptive epidemiology of health state values and utilities. Med Decision Making 1993;13:87-8.

55. Clancy C, Cebul R, Williams S. Guiding individual decisions. A randomized, control trial of decision analysis. Am J Med 1988;84:283-8.

56. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. Health Serv Res 1976;11:478-507.

57. Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preference for health states. Operations Res 1982;30:1043-69.

58. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health classification systems: Health Utilities Index. PharmacoEconomics 1995;7:490-502.

59. Boyle MH, Furlong W, Feeny D, Torrance GW, Hatcher J. Reliability of the Health Utilities Index-Mark III used in the 1991 cycle 6 General Social Survey Health Questionnaire. Qual Life Res 1995;4:249-57.

60. The EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.

61. Rosser R, Cottee M, Rabin R, Selai C. Index of health-related quality of life. In: Hopkins A, ed. Measures of the quality of life and the uses to which such measures may be put. London: Royal College of Physicians of London, 1992:81-90.

62. Brazier J, Jones N, Kind P. Testing the validity of the EuroQol and comparing it with the SF-36 health survey questionnaire. Qual Life Res 1993;2:169-80.

63. Anderson RT, Aaronson NK, Wilkin D. Critical review of the international assessments of health-related quality of life. Qual Life Res 1993;2:369-95.

64. Essink-Bot ML, Stouthard MEA, Bonsel GJ. Generalizability of valuations on health states collected with the EuroQol-questionnaire. Health Econ 1993;2:237-46.

65. Rosser R, Sintonen H. The EuroQol quality of life project. In: Walker SR, Rosser RM, eds. Quality of life assessment: key issues in the 1990s. London: Kluwer Academic Publishers, 1993:197-9.

66. Rosser R, Allison R, Butler C, Cottee M, Rabin R, Selai C. The Index of Health-Related Quality of Life (IHQL): a new tool for audit and cost-per-QALY analysis. In: Walker SR, Rosser RM, eds. Quality of life assessment: key issues in the 1990s. London: Kluwer Academic Publishers, 1993:179-84.

67. Detsky AS, Naglie IG. A clinician's guide to cost-effectiveness analysis. Ann Intern Med 1990;113:147-54.

68. Mason J, Drummond M, Torrance G. Some guidelines on the use of cost effectiveness league tables. BMJ 1993;306:570-2.

69. Petrou S, Malek M, Davey PG. The reliability of cost-utility estimates in cost-per-QALY league tables. PharmacoEconomics 1993;3:345-53.

70. Hornberger JC, Redelmeier DA, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. J Clin Epidemiol 1992;45:505-12.

71. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. Med Care 1981;19:787-805.

72. Boyle MH, Torrance GW, Sinclair JC, Horwood SP. Economic evaluation of neonatal intensive care of very-low-birthweight infants. N Engl J Med 1983;308:1330-7.

73. Oldridge N, Furlong W, Feeny D, Torrance G, Guyatt G, Crowe J, et al. Economic evaluation of cardiac rehabilitation soon after acute myocardial infarction. Am J Cardiol 1993;72:154-61.

74. Cook J, Richardson J, Street A. A cost utility analysis of treatment options for gallstone disease: methodological issues and results. Health Econ 1994;3:157-68.

75. Goodwin PJ, Feld R, Evans WK, Pater J. Cost-effectiveness of cancer chemotherapy: an economic evaluation of a randomized trial in small cell lung cancer. J Clin Oncol 1988; 6:1537-47.

76. Smith TJ, Hillner BE. The efficacy and cost-effectiveness of adjuvant therapy of early breast cancer in premenopausal women. J Clin Oncol 1993;11:771-6.

77. Smith RD, Hall J, Gurney H, Harnett PR. A cost-utility approach to the use of 5-fluorouracil and levamisole as adjuvant chemotherapy for Dukes' C colonic carcinoma. Med J Aust 1993;158:319-22.