**Lecture 6 Dr. Haider Raheem**

**Cost-Utility Analysis**

**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 cost-effectiveness analysis in which the measure of effectiveness is a utility- or preference-adjusted outcome.

**Overview**

As mentioned in lecture 3, some consider **cost-utility analysis** (CUA) a subset of CEA because the outcomes are assessed using a special type of clinical outcome measure, usually the **quality-adjusted** **life-year** (QALY). A CUA takes patient preferences, also referred to as **utilities,** into account when measuring health consequences. Some authors prefer to use other terms, such as *preference weight* or *preference value*, in place of the word *utility*. The most common outcome unit used in CUA is the QALY, which incorporates both the quality (morbidity) and quantity (mortality) of life. Other outcome units that are seen less frequently include disability-adjusted life-years (DALYs) and healthy-year equivalents (HYEs), among others.

 The advantage of a CUA is that different types of health outcomes and diseases with multiple outcomes of interest can be compared (unlike in CEA) using one common unit such as the QALY. CUA incorporates morbidity and mortality into this one common unit without having to determine or estimate the monetary value of these health outcomes (unlike CBA). The disadvantage of this method is that it is difficult to determine an accurate utility or preference weight value. Therefore, although the number of CUA research articles in the literature is increasing yearly, the methods for estimating utilities/QALYs may not be fully understood or embraced yet by many US providers or decision makers.

 For some research questions, utility adjustments may not be warranted. For example, if two pharmaceutical products have different outcomes based on the number of life-years saved (LYS), but the quality of each year of life for those on the two treatments are thought to be the very similar (see Fig. 6.1A), quality adjustment may not be as crucial. However, in many cases—for example, cancer treatment—both the length of life and the quality of life are different, depending on the therapy selected. Sometimes the treatments that extend life the longest are also the most toxic, so a measure that incorporates both length of life and quality of life is needed in these cases (see Fig. 6.1B). Many health conditions do not have an impact on patients’ length of life, but *only* on the quality of their life, and CUA may be a good choice for comparing treatments for these conditions (see Fig. 6.1C). Examples include conditions such as hearing loss, seasonal allergies, and erectile dysfunction. CUAs may also be useful when comparing treatments and outcomes that are very different (e.g., when comparing the treatment of heart disease with prenatal care) because outcomes for both treatments can be summarized into one common unit, such as QALYs.

 By convention, perfect health is assigned a value of 1.0 utility (*μ*) per year, and death is assigned a value of 0.0. If a person’s health is diminished by disease or treatment, 1 year of life in this state is valued somewhere between 0 and 1. Some researches point out that there are disease states worse than death, so negative utility weights may be needed to depict these values. In the vast majority of studies, this is not an issue, and the values for each year are estimated to be between 0.0 and 1.0.

 To estimate utility weights for various conditions or “health states” between perfect health and death, two broad methods are used to elicit, or generate, these scores: direct elicitation and indirect elicitation. Direct elicitation methods (**rating scale**, **standard gamble**, and **time tradeoff**). Indirect elicitation methods, using standardized weightings (e.g., **EQ-5D** and **SF-6D** surveys).

**When Is Cost-Utility Analysis Appropriate?**

Drummond *et al*.enumerated several circumstances in which CUA may be the most appropriate analytic approach:

1. When health-related 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 (eg, treatments for osteoarthritis).

2. When health-related 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 (eg, the impact of a treatment-induced stroke in a survivor).

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 (eg, heart disease), but may increase mortality from other conditions (eg, 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.



**Fig. 6.1 Examples estimating and comparing quality-adjusted life-years (QALYs). A:** Example illustrating when two treatment options produce different outcomes on the basis of the number of life years saved (LYS), but the quality of each year of life for those on the two treatments are thought to be the very similar (quality adjustment not needed). **B:** Example illustrating when both the length of life and the quality of life are different depending on the therapy selected (quality adjustment appropriate). **C:** Example illustrating options that do not have an impact on patients' length of life but only on the quality of their life (quality adjustment appropriate).

**Steps in Calculating QALYs**

To calculate QALYs, the following steps apply:

1. Develop a description of each disease state or condition of interest.

2. Choose a method for determining utilities.

3. Choose subjects who will determine utilities.

4. Sum the product of utility scores by the length of life for each option to obtain QALYs.

Each of these steps is explained below.

**Step 1: Develop a Description of Each Disease State or Condition of Interest**

The description should concisely depict the usual health effects expected from the disease state or condition. It should include the amount of pain or discomfort, any restrictions on activities, the time it may take for treatment, possible changes in health perceptions (worry or concern), and any mental changes. Example describing hospital-based kidney dialysis is presented below:

***Description of Hospital-Based Kidney Dialysis***

You often feel tired and sluggish. A piece of tubing has been inserted into your arm or leg, which may restrict your movement. There is no severe pain but rather chronic discomfort. You must go to the hospital twice a week for 6 hours per visit. You must follow a strict diet (low salt, little meat, no alcohol). Many people become depressed because of the nuisances and restrictions, and some feel they are being kept alive by a machine.

**Step 2: Choose Method for Determining Utilities**

The three most common methods for determining preference, or utility, weights are rating scales (RS), standard gamble (SG), and time tradeoff (TTO). For each of these methods, a disease state or condition or multiple disease states or conditions are described to subjects who help determine where these disease states or conditions fall between 0.0 (dead) and 1.0 (perfect health).

***Rating Scale***

An RS consists of a line on a page with scaled markings, somewhat like a thermometer with perfect health at the top (100) and death at bottom (0). An instrument called the Visual Analog Scale (VAS) is similar to the RS, but it does not have any markings between the best and worst scores, and subjects are told to mark an “X” somewhere between the two extremes to indicate their preferences. As an example, if they place a disease state at 70 on the scale, the disease state is given a utility score of 0.7. Most people would agree that mild seasonal allergies would not decrease a person’s quality of life as much as being in a coma for the year. Therefore, the preference score for mild allergies would be near the 1.0 (or 100) mark at the top of the RS, and the value for being in a coma would be near 0, or the bottom of the scale.



**Fig. 6.2 Rating scale (RS) with example estimates for various disease states or conditions.** The RS uses a thermometer-like illustration to ask respondents to estimate the utility of different health states ranging from 0 (dead) to 1.0 (or 100; perfect health). In this example, the respondent estimated that being in a coma for 1 year has a lower utility (0.1) than having hospital dialysis for 1 year (0.6). Both are lower than the utility estimate for mild allergies (0.9).

***Standard Gamble***

For this method, each subject is offered two alternatives. Alternative 1 is treatment with two possible outcomes: either the return to normal health or immediate death. Alternative 2 is the certain outcome of a chronic disease state for life based on a person’s life expectancy (Fig. 6.3). The probability, or *p*, of normal health (versus immediate death, or 1− *p*) for Alternative 1 is varied until the subject is indifferent between Alternatives 1 and 2 (living with the disease state or condition).



**Fig. 6.3 Standard gamble (SG).** Using the SG approach, the respondent is asked to think about being in a chronic health state and then told that he or she could gamble on an intervention (e.g., an operation) that could either cure the condition (probability = *p*), although he or she might die from the intervention (probability = 1- *p*). A base probability is given and the respondent is asked whether he or she would have the intervention or live with the chronic condition. This probability is varied until the respondent is indifferent (the two options are difficult to choose between). The probability at this indifference point is the utility associated with the condition.

 As an example, a person considers two options: a kidney transplant with a 20% probability of dying (80% chance of returning to normal health) during the operation (Alternative 1) or certain dialysis for the rest of his or her life (Alternative 2). If the person says he or she would have the operation if the chance of the successful operation *p* is 80% (chance of immediate death, 20%), the percent chance of success is *decreased* until the person reaches his or her point of indifference (the point where the two options are nearly equal and the person cannot decide between the two). If the person says he or she would not have the operation if the percent chance for success was 80% (chance of dying, 20%), the percent chance of success is *increased* until the person reaches his or her point of indifference. Let us say that the first person chooses a 70% chance (*p*) of a successful operation (with a 30% chance [1− *p*] of immediate death) as the point of indifference between having a kidney transplant and living with kidney dialysis for life. The utility score for this person for this disease state or condition (kidney dialysis) would be calculated as the probability (*p*) of living a normal life after the operation, or 0.7.

***Time Tradeoff***

The third technique for measuring health preferences, or utilities, is the TTO method (Fig. 6.4). Again, the subject is offered two alternatives. Alternative 1 is a certain disease state for a specific length of time (*t*), the life expectancy for a person with the disease, and then death. Alternative 2 is being healthy for time *x*, which is less than *t*. Time *x* is varied until the respondent is indifferent between the two alternatives. The utility score for the health state is calculated as *x* divided by *t*.

****

**Fig. 6.4 Time tradeoff (TTO).** This TTO schematic represents the choice a respondent makes about trading off years of life for better health for a shorter period of time. The respondent is given the choice of living a full life (to time *t*) with a specific condition or living fewer years (to time *x*) without the condition (being healthy). The time of living healthy is varied until the respondent is indifferent between living in full health *x* years and living with the condition for *t* years. The utility calculated for the condition is *x*/*t*.

 For example, a person with a life expectancy of 40 years is given two options: Alternative 1 is having a chronic condition (e.g., kidney disease or diabetes) for 40 years, and Alternative 2 is being healthy (no disease) for 20 years followed by death. If the person says he or she would rather have the disease for 40 years (*t*) than be healthy for 20 years, the number of years (*x*) in the healthy state is *increased* until the person is indifferent between the two alternatives. If the person would rather be healthy for 20 years than have the disease for 40 years, the number of years (*x*) in the healthy state is *decreased* until the person is indifferent between the two alternatives. Let us say that for a person who expects to live 40 more years, the person’s point of indifference is 30 years of health versus 40 years of kidney disease. The utility score would be *x*/ *t* = 30 / 40 or 0.75. As with the SG method illustrated above, these calculations are for chronic diseases or conditions, and calculations for a temporary health state are more complex and can be found elsewhere.

**Comparisons of the Three Methods**

The advantage of using the RS method to determine utilities is that many disease states or conditions can be described to each subject, and this method can be conducted via a questionnaire without face-to-face interaction. One disadvantage of using the RS method is that it does not incorporate time into the utility score as easily as the other two methods.

 The advantage of using the SG method is that it is the “gold standard” and based on economic theory. It is more difficult for the participants, and few disease states or conditions can be “cured” by an intervention that brings a person back to “normal health.” This is better administered in a face-to-face setting, or through an iterative process.

 Some advantages of the TTO method are that it is more adaptable to diseases states than the SG, and it incorporates the time in the disease state or condition more easily than the RS.

 Unfortunately, the average utility scores for each disease state or condition may differ depending on which method is used. RS scores have been shown to be consistently lower than either SG or TTO scores, and TTO scores are sometimes lower than SG scores.

**Step 3: Choose Subjects Who Will Determine Utilities**

In the previous examples of the three methods, the term *subject* was used to describe the person who would be questioned to determine the utility, or preference scores. Who is this subject? Who should determine utilities, the patient with the disease, the health care professional, the caregiver, or people from the general public?

 An advantage of eliciting utility scores from patients with the disease or condition of interest is that these patients may understand the effects of the disease better than the general public. However, some believe these patients provide a biased view of their disease compared with other diseases. In many cases, a patient with the specific disease state or condition reports higher utility scores than others (e.g., general population, caregivers).

 Some contend that health care professionals could provide good estimates because they understand various diseases. Others argue that these professionals may not rate discomfort and disability as seriously as patients or the general public. For example, researchers found that when patients were asked about side effects of medications to treat hepatitis C, patients were more concerned (gave lower utility scores) than the providers about the side effects of treatment.

 In the literature, health care professionals are often asked to determine utility scores. This may be based on practicality because these professionals have had experience with the disease states and are easily accessible for interviews.

**Step 4: Multiply Utilities by the Length of Life for Each Option to Obtain QALYs**

When comparing the options, the difference in the length of life permitted by each option is multiplied by the utility scores obtained above. For example, in Table 6.1, we will assume that the utility score for each year of additional life is constant. In actuality, for many conditions, the utilities would change over time as the condition improves or worsens (see Fig. 6.1B).

 In Table 6.1, we compare two treatment options, drug A and drug B. Although drug B extends the person’s life for more years, the quality of life for those years is lower than with drug A. If a CEA were conducted, option B would be relatively cost-effective at an incremental cost per year of life of $5,000. If the quality of those years is incorporated into the equation by calculating QALYs, option A becomes **dominant** in that it costs less and provides a better outcome (more QALYs).

Table 6.1 QALY Calculations.



**Cost-Utility Analysis and Healthcare Interventions**

Table 6.2 illustrate the different elements that are needed to conduct CEA and CUA when comparing two alternatives. An assumption in the table is that quality of life (utility) remains constant over the full life expectancy.

Table 6.2 Economic analysis of two alternatives treatment interventions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Intervention** | **Cost ($)** | **Effectiveness (life expectancy)** | **Health state (utility)** | **QALYs** |
| Treatment A | 20,000 | 4.5 | 0.60 | 2.7 |
| Treatment B | 10,000 | 3.5 | 0.72 | 2.5 |
| Incremental cost-effectiveness ratio$=\frac{\$20,000-\$10,000}{4.5 y-3.5 y}$ = $10,000 per life-year gained |
| Incremental cost-effectiveness ratio$=\frac{\$20,000-\$10,000}{2.7 QALY-2.5 QALY}$ = $50,000 per QALY gained |

**Incremental Net Benefit Analysis**

Oncoplatin cost $3,000 more than Oncotaxel ($10,000 versus $7,000, respectively) and produced an additional 0.04 QALY (0.19–0.15); therefore, the incremental ratio was $75,000 per extra QALY. Is Oncoplatin more cost-effective than Oncotaxel? It depends on the value of a QALY. There has been debate on the value of a QALY, and broad ranges of estimates have appeared in the literature. The most often-cited value from the US literature is $50,000 per QALY. Below are incremental net benefit calculations using a λ of $50,000 as an estimate of the value of the health benefit of one QALY.

 INB = (λ × Δ QALYs) − Δ Costs

 INBλ=$50,000 = ($50,000 × 0.04 QALY) − $3,000

 INBλ=$50,000 = −$1,000

Because the INB is less than zero, Oncoplatin is not cost-effective compared with Oncotaxel when λ = $50,000. The $50,000 per QALY estimate has been cited since the 1980s without adjustment for inflation and would be approximately double if inflated to 2012 dollars. The figure that follows indicates that if the λ were varied from $20,000 to $100,000 per QALY, the INB of Oncoplatin compared with Oncotaxel would range from −$2,200 to more than $1,000, indicating that the answer to “Is Oncoplatin cost-effective compared with Oncotaxel?” would depend on (i.e., be sensitive to) the value placed on a QALY (λ).

****

**Fig. 6.5 Calculation of INB for range of lambda values for QALYs.**

**Health-Related Quality of Life**

**Definitions**

It may be helpful to distinguish between the terms **quality of life** (QoL) and **health-related quality of life (HRQoL)**. The first term, QoL, is a broad concept with many aspects that measures people’s overall perception of their life. QoL includes both health-related and non–health-related aspects of their lives (e.g., economical, political, cultural). HRQoL is the part of a person’s overall QoL that “represents the functional effect of an illness and its consequent therapy upon a patient, as perceived by the patient.”

**HRQoL Measures Vs. Utility Measures**

Table 6.3 Comparison of health measures.



**General or Generic Measures**

The advantage of **generic health status** instruments is that scores can be compared for many disease states and conditions. On the other hand, general measures may not be sensitive to clinically relevant differences for every disease or condition.

Table 6.4 Examples of General HRQoL measures.

****

***The Medical Outcomes Short-Form Surveys***

The SF-36 (the short form of the Medical Outcomes Survey that consists of 36 items) is a multipurpose survey of general or generic health status. The SF-36 yields a profile of eight concepts as well as summary physical and mental health measures. The SF-36 also includes a self-evaluation of change in health during the past year. Both standard (4-week) and acute (1-week) recall versions have been published.

 A slightly different newer version (version 2) and a shorter version consisting of 12 questions (the SF-12) have also been developed. The goal of creating the SF-12 was to be brief enough for practical use yet encompass the important physical and mental measures of the full SF-36.

**Domains of Health Status**

Four essential dimensions, or domains, should be included in all HRQoL instruments: physical functioning, psychological functioning, social and role functioning, and general health perceptions.

**Assessing HRQoL Instruments**

Similar to utility measures, assessments using HRQoL measures may be viewed cautiously by prescribers or decision makers who have been trained to collect and interpret “hard, objective” data, such as blood pressure measurements, radiography results, and blood concentrations of biologic markers. Patient-based assessments of pain, depression, or anxiety are examples of important outcomes that are more subjective in nature. Therefore, it is important to assess the psychometric properties of these HRQoL instruments, including consistency (i.e., **reliability**), precision (i.e., **validity**), and their ability to measure meaningful clinical changes (i.e., **responsiveness**).

**Reliability**

Reliability refers to the consistency of an instrument. For example, does the instrument produce the same score on multiple administrations? The three common types of reliability measures are test–retest reliability, internal consistency, and interrater reliability.

 Note that an instrument can be tested and found to be **reliable** (i.e., it elicits similar scores on readministration) yet imprecise (not **valid**). In other words, scores calculated using an HRQoL instrument might be reliably (consistently) wrong. If a physician’s office had a scale to measure patients’ weights, and this scale indicated a weight for each patient that was exactly 25 lb lower than each person’s actual weight, this scale would be considered to be reliable (consistent from visit to visit) but not valid (wrong estimation of patients’ true weights). Figure 6.6 illustrates the difference between the terms reliability and validity.



**Fig. 6.6.** This figure uses three targets to indicate how accurate the scores (*bullets*) are to the real measure of the domain or construct (*bulls-eye* or *center of target*). The left target illustrates scores that are neither reliable (consistent) nor valid (precise). The middle target illustrates the concept of scores that are reliable but not valid. The bullets tend to hit the same area every time, but they are not near the center of the target. The right target illustrates scores that are both reliable and valid—the score is consistently near the center of the target.

**Validity**

Validity studies are necessary to evaluate whether the scores elicited from the instruments truly represent the underlying constructs (aspects) of HRQoL. In other words, the purpose of validity assessment is to determine whether the instrument is actually measuring what it is supposed to be measuring. The validity of an instrument is much more difficult to assess than its reliability.

 When enough evidence has been accumulated to show that an instrument reflects the health concept intended to be measured and that it does not measure other unintended concepts, researchers say that a HRQoL instrument has been “validated.”

**Responsiveness**

The responsiveness of an HRQoL instrument refers to its ability to detect changes in health status. This includes the instrument’s ability to not only show numerical differences in scores between patients in dissimilar health states but to also detect changes in a patient’s health over time when his or her health status changes.