Historically, surgical outcomes have been measured by rates of mortality and morbidity. Fortunately, in plastic surgery, mortality is rarely an issue unless one is practicing in high-risk settings such as burn units or war zones. The improvement of presurgical morbidity, however, is very relevant to our specialty, and great effort should be made to measure it as precisely as possible. The success of plastic surgery interventions has traditionally been measured by before-and-after photographs in cosmetic surgery, pinch and grip strength and range of motion in hand surgery, percentage of successful flaps or replanted digits in microsurgery, and so forth. Although such outcomes are important, they serve only as proxies for the ultimate goal of optimal physical, mental, and social well-being of the patient. Would one seriously consider a face-lift result satisfactory if the patient required psychiatric support after her operation despite a much improved postoperative photograph? Likewise, would an improvement of 0.5 kg in grip strength after a novel hand procedure be considered successful if the patient was still unable to return to work? What about a meticulously executed flap to reconstruct a through-and-through defect of the mandible in a head and neck cancer where the patient died 6 months later from cancer recurrence? Plastic surgeons have recognized for some time now that there can be a discrepancy between their interpretation of a successful surgical outcome and the patient’s expectation.

A rapid expansion of outcomes research has occurred over the past two decades in many medical specialties. This has been fueled, at least in part, by the practice of evidence-based medicine. Evidence-based medicine integrates individual clinical expertise with the best available clinical evidence from systematic research. Outcomes research sets out to measure a patient’s quality of life. The broader term “quality of life” can be defined as “the adequacy of people’s material circumstances and their feelings about these circumstances.” This encompasses indicators of life satisfaction; personal wealth and possessions; level of safety; level of freedom; spirituality; health perceptions; and physical, psychological, social, and cognitive well-being. Health-related quality of life, a subcomponent of quality of life, includes all areas specific to health, that is, physical, emotional, psychological, social, cognitive, and role functioning, in addition to abilities, relationships, perceptions, life satisfaction, and well-being. It refers to patients’ appraisals of their current level of functioning and satisfaction with it, compared with what they perceive to be ideal. There are several different methods available to measure a patient’s health-related quality of life. In the past two decades, a number of validated and reliable scales have been introduced to measure improvement in quality of life of plastic surgical interventions.

One particular health-related quality-of-life outcome measure, although infrequently used in plastic surgery, is the quality-adjusted life-year. The quality-adjusted life-year was originally introduced as a measure of health effectiveness for cost-effectiveness analysis to assist decision makers in allocating scarce resources across competing health care programs. Health care spending in the United States is the highest in the world, averaging $7026 per person per year. However, the U.S. health care system ranks 37th among the 191 countries. Although nationwide efforts to decrease this unsustainable growth have had limited success so far, this issue will be dealt with sooner
The most rational way for present and future governments to control this growth would be to demand financial accountability for our surgical interventions, as measuring the effectiveness alone is not enough. It is especially important to consider the cost of new technologies, which is often greater than the cost of prevailing ones. In some jurisdictions, for certain medical interventions, financial accountability is the norm. The proper methodology for integrating both the effectiveness and associated costs is through a cost-effectiveness analysis or its variant, cost-utility analysis. As the quality-adjusted life-year is an important component of a cost-utility analysis, use of this outcome measure will propel plastic surgery to the forefront of those specialties that adopt evidence-based clinical practice.

The purpose of this article is to familiarize plastic surgeons with the quality-adjusted life-year as a surgical outcome measure. This article explains what a quality-adjusted life-year is and how it is measured, the concept of the quality-adjusted life-year, and how to use quality-adjusted life-years in plastic surgery.

WHAT IS A QUALITY-ADJUSTED LIFE-YEAR AND HOW IS IT MEASURED?

The quality-adjusted life-year estimates the length and quality of extra years gained by a medical/surgical intervention. The concept is based on the experience of an individual’s health state over time, and each health state has a value attached to it. Health states have a value on a scale of 0.0 to 1.0, with 0.0 being dead and 1.0 being perfect health. An increasing value is equated with an increasing preference for a particular health state. This preference is referred to as a utility.

Utilities can be obtained from quality-of-life assessment instruments, including visual analogue scales, the standard gamble, time-trade off, and multiattribute scales.

Quality-of-life assessment instruments are surveys or questionnaires that ask the respondent about his or her own current health state. Visual analogue scales or “feeling thermometers” typically ask individuals to rate their preference on a scale of 0 to 1, with 1 being the best possible outcome. The scale is typically 100 mm in length, anchored by word descriptors at each end. Visual analogue scales are advantageous in that they can be administered quickly and easily. However, they only measure one health state.

The standard gamble is based on the third axiom (continuity of preference) of the utility theory by von Neumann and Morgenstern. If there are three outcomes such that $X_1$ is preferred to $X_2$, which is preferred to $X_3$ ($X_1 > X_2 > X_3$) there is some probability $p$ at which the individual is indifferent between outcome $X_2$ with certainty or receiving the risky prospect made up of outcome $X_1$ with probability $p$ and outcome $X_3$ with probability $1 - p$. This method has been used extensively in decision analysis, and can be used to measure preferences for chronic conditions. The standard gamble requires respondents to answer questions based on a hypothetical scenario consisting of two choices (A and B). In general, respondents are reporting on a health state different from their own current health state. Cugno et al. used the standard gamble and time-trade-off techniques to calculate the quality-adjusted life-years gained by composite tissue allotransplantation for patients with severe facial disfigurement. Using the standard gamble, choice A was to undergo the risky surgical procedure, composite tissue allotransplantation, and choice B was to remain in the current health state of severe facial disfigurement. Choice A would be considered a “gamble,” as the outcomes are unknown and vary in probabilities of best possible outcome (i.e., successful surgery without complications), associated complications (e.g., acute rejection, graft loss, diabetes), and worst possible outcome (i.e., immediate death). Participants were asked to estimate the largest percentage risk of complications (or death) that they would be willing to ac-
cept to be relieved of their current health state (severe facial disfigurement). The health states and their utilities, transformed into quality-adjusted life-years, are illustrated in Figure 3. The quality-adjusted life-year associated with severe facial disfigurement was 9.36 years. Standard gamble is theoretically sound but involves complex measurements, which may not be easy for respondents. This technique may also be subject to cognitive biases. In 1982, Torrance and colleagues developed the time trade-off as an alternative to the standard gamble.21

Time trade-off (Fig. 4) is similar to standard gamble in that respondents answer questions based on hypothetical scenarios. Using time trade-off, choice A is a health state (i.e., severe facial disfigurement) associated with a life expectancy of $t$ years. Choice B is perfect health with a shortened

**Fig. 2.** The standard gamble.

**Fig. 3.** Decision analytic tree for composite tissue allotransplantation of the face, including adjusted probabilities and quality-adjusted life-years for each health state (pathway). QALYs, quality-adjusted life-years; PTLD, posttransplant lymphoproliferative disease; CMV, cytomegalovirus. (From Cugno S, Sprague S, Duku E, Thoma A. Composite tissue allotransplantation of the face: Decision analysis model. Can J Plast Surg. 2007;15:145–152.)
life expectancy of $x$ years (composite tissue allotransplantation with no complications). Time in perfect health ($x$) is varied to find the value ($t$), where choice A is equal to choice B. The utility value of the health state in question is estimated by the proportion between the shortened life expectancy $x$ and the full life expectancy of the intermediate outcome $t$. In other words, the respondent is asked how much time (years of life) he or she would be willing to give up being in a healthier state compared with a less healthy state. If a respondent is willing to give up one-tenth of his or her life expectancy in return for perfect health, his or her utility for the intermediate health state is 0.1. Cugno et al.20 found the quality-adjusted life-years for severe facial disfigurement to be 8.9 using the time trade-off technique. With composite tissue allotransplantation, there is an expected gain of 16.2 quality-adjusted life-years.

The main controversy surrounding the standard gamble and time trade-off is whose perspective is most important (i.e., medical professionals familiar with the illness and outcomes, the general public who do not have the illness, or patients suffering from the illness)?22 Often, the respondents to the standard gamble and time trade-off are medical professionals; thus, their responses may not be generalizable. Responses from the general public may not actually reflect those of patients suffering from the illness. In general, the Panel on Cost-Effectiveness in Health and Medicine advocates that the perspective of the general public (society) be taken.22,23 Depending on the clinical problem, it may be difficult to collect responses from patients suffering from the illness, for example, children, the mentally ill, and those with rare conditions/procedures (i.e., composite tissue allotransplantation patients). The general public should be a representative sample including patients and disabled people as they occur. They must be informed of the health states involved. It should be noted that this technique may be prone to bias, depending on how the health state is described to the subject, who may or may not comprehend thoroughly the clinical problem.

Multiattribute scales are the preferred method of obtaining utilities by the Panel on Cost-Effectiveness in Health and Medicine and the National Institute of Health and Clinical Excellence.22,23 These scales combine several health attributes (domains) that can be used to calculate utilities. The most commonly used multiattribute scales are listed in Table 1.24–30 Multiattribute scales provide an overall assessment of quality of life; however, they may not be sensitive enough to detect small differences in patient groups with specific disabilities.5 In instances where cost-effectiveness analysis is not the primary objective, one may use disease-specific scales that consist of questions focusing on specific symptoms and impairments relevant to a particular disease state or surgical intervention.5 A recommendation was made by Guyatt et al.31 to include a generic instrument, a disease (condition)-specific instrument, and a utility measurement in the evaluation of medical interventions whenever possible. The difference in quality of life obtained from multiattribute scales should be reported and interpreted in a clinically meaningful way.5,32,33 For example, what does a six-point difference in the Short Form-36 mean? Is this change in score small, medium, or large? It has been suggested as a rule of thumb that a 10-point change in scores on 100-point quality-of-life scales such as the Short Form-36 is clinically
With regard to utilities, the minimal clinically important difference in Health Utilities Index is 0.03. Which technique is used (i.e., visual analogue scale, standard gamble, time trade-off, multiattribute scales) will depend on what the question is and whom we ask. These questions could pertain to individual or community health. For most plastic surgery interventions, it is possible to obtain utilities directly from patients. An example would be in an ongoing randomized controlled trial comparing the vertical scar to the inferior pedicle reduction mammaplasty. Utility measurements are being obtained using the Health Utility Index Mark2/3 instrument. Quality-of-life measurement is also obtained through the Short Form-36 (generic scale) and the Breast-Related Symptom Questionnaire (disease-specific) scales. If we want to know whether composite tissue allotransplantation of the face should be covered by some national health care system (i.e., Medicare), we need to obtain utilities from the general public, as the society has a vested interest in the allocation of health care resources. Table 2 shows the techniques that can be used depending on the question asked and the perspective taken.

### THE CONCEPT OF THE QUALITY-ADJUSTED LIFE-YEAR

Utilities are qualitative measures of quality of life (qualitative gains). In measuring the benefit of a surgical intervention, however, we also need to specify the gains in reduced mortality (quantitative gains). The concept of the quality-adjusted life-year can be explained using the three diagrams in Figures 5 through 7. Suppose in a randomized controlled trial we compare two fluid resuscitation protocols on severely burned patients. We will call these the “novel” and the “standard” approaches. The novel approach provides a mean utility of 0.6 compared with the standard approach with a mean utility of 0.5 (for a moment, let us set aside the issue on how we could measure the utility in such patients). With the novel resuscitation approach, we find out that the mean gain in life is 2 years compared with 1 year with the old approach. If we were to plot utilities and time (Fig. 5), the areas under the curves would represent the different gains.

### Table 1. Multiattribute Utility Scales

<table>
<thead>
<tr>
<th>Name of Scale</th>
<th>Domains</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Utilities Index Mark 2</td>
<td>Sensation (four levels), mobility (four levels), emotion (five levels), cognition (four levels), self-care (four levels), pain (five levels), and fertility (three levels).</td>
<td>The scores fall on the 0.0 (dead) to 1.0 (perfect health) value scale.</td>
</tr>
<tr>
<td>Health Utilities Index Mark 3</td>
<td>Vision (six levels), hearing (six levels), speech (five levels), ambulation (six levels), dexterity (six levels), emotion (five levels), cognition (six levels), and pain (five levels).</td>
<td>The scores fall on the 0.0 (dead) to 1.0 (perfect health) value scale.</td>
</tr>
<tr>
<td>European Quality of Life-5 Dimensions (EQ-5D)</td>
<td>Mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each attribute comprises three levels.</td>
<td>The scores fall on the 0.0 (dead) to 1.0 (perfect health) value scale.</td>
</tr>
<tr>
<td>Short Form-6D (SF-6D)</td>
<td>Physical functioning (six levels), role limitations (four levels), social functioning (five levels), pain (six levels), mental health (five levels), and vitality (four levels).</td>
<td>To use Short Form-6D, you must use the Short Form-36 or Short Form-12 plus three additional questions to collect the data to classify the patients into the Short Form-6D. Data are put into a scoring table to calculate utilities.</td>
</tr>
<tr>
<td>Quality of Well-Being (QWB)</td>
<td>Mobility, physical activity, social activity, and symptom-problem complex.</td>
<td>Respondents rank health states on a scale of the 0.0 (dead) to 1.0 (perfect health). Scores are values, not utilities.</td>
</tr>
</tbody>
</table>

### Table 2. Quality-Adjusted Life-Years: Matrix of Perspective and Valuation Techniques*

<table>
<thead>
<tr>
<th>Question</th>
<th>Personal Clinical Decision</th>
<th>Societal Audit: Evaluation of Ongoing Surgical Activities (Practices)</th>
<th>Societal Resource Allocation: Priority Setting across Proposed Surgical Programs or Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whom to ask</td>
<td>The individual</td>
<td>Those affected by the surgical interventions (e.g., patients)</td>
<td>Representative sample of the population</td>
</tr>
<tr>
<td>Valuation techniques</td>
<td>SG, TTO, VAS</td>
<td>SG, TTO, or MAU instrument</td>
<td>SG, TTO, VAS, or MAU instrument</td>
</tr>
</tbody>
</table>

SG, standard gamble; TTO, time trade-off; VAS, visual analogue scale; MAU, multiattribute utility.

*Adapted from Weinstein MC, Torrance G, McGuire A. QALYs: The basics. Health Serv Res. 2009;12:S5–S9.)
sent the gains of each approach. The difference in the areas is the gain of the novel resuscitation approach. Mathematically, a quality-adjusted life-year is calculated using the following formula:

\[
\text{quality-adjusted life-year} = \frac{\text{duration of health state}}{\text{utility of health state}} \times \frac{\text{future remaining life expectancy} - \text{duration of health state}}{\text{utility of best possible outcome}}.
\]

Simplifying the hypothetical fluid resuscitation example cited above, the quality-adjusted life-years gained would be as follows: quality-adjusted life-years gained = quality-adjusted life-years “novel” approach (2 × 0.6) − quality-adjusted life-years “standard” approach (1 × 0.5) = 0.7 quality-adjusted life-year.

How do we interpret this value of 0.7 quality-adjusted life-year gained? In simple terms, it tells us that the new resuscitation approach prolongs the life of the “burned patient” by 0.7 year in perfect health.

Let us consider now the scenario where a new surgical intervention may prolong the life of a patient (quantity) but not its quality. Theoretically, such a patient could be a comatose patient with pressures sores on a chronic hospital ward. It is conceivable that by performing surgery (excision of the ulcers and closure with myocutaneous flaps) compared with doing nothing we could prolong the life of the patient by \(x\) years, as he or she will not die as a result of sepsis or complications of

Fig. 5. The “novel” approach improves quantity of life but has a smaller effect on quality of life.

Fig. 6. Surgery improves the quantity of life but does not substantially affect quality of life.
sepsis. The quality of life may not be substantially altered by the procedure (Fig. 6).

Most plastic surgical interventions are more likely to be represented by Figure 7. It is highly unlikely that a novel cosmetic intervention or a new breast reconstruction method will prolong the life of the patient beyond year $x$, but it is plausible that it will improve the quality of life (at least this is what we hope for).

As many plastic surgical interventions do not have death under consideration, one may suspect that the utilities will typically fall in the range of 0.9 to 1.0. As a result, one may question the precision of the quality-adjusted life-year in such a tight range. We believe existing utility scales are sensitive to capture the observed changes in most plastic surgical interventions. For example, in a prospective cohort study of 50 consecutive breast reduction patients,\textsuperscript{35} the authors found a 0.13 gain in utility from the day before surgery to 1 year after surgery using the Health Utility Index Mark2/3. The Health Utility Index Mark2/3 (Table 1) has the ability to identify close to 1 million unique health states. Before surgery, patients experienced on average a utility of 0.76. After surgery, patients experienced on average a utility of 0.88, a gain of 0.12 quality-adjusted life-years (Fig. 8). Therefore, the quality-adjusted life-years gained by breast reduction surgery would be as follows: quality-adjusted life-years = 1 year $\times$ 0.12 + 40 years $\times$ 0.13 = 5.32 quality-adjusted life-years (average age was 38, assuming life span of 79 years). This is equivalent to a patient living an additional 5.32 years in perfect health. This in essence is the benefit of breast reduction surgery. The same methodology can be used to compare, for example, the minimal access cranial suspension lift and the classic face lift or the superficial inferior epigastric artery-to-deep inferior epigastric perforator flap breast reconstruction after mastectomy.

**USING QUALITY-ADJUSTED LIFE-YEARS IN PLASTIC SURGERY RESEARCH**

The quality-adjusted life-year as an outcome measure can be used to compare interventions across different patient populations, making it applicable to all areas of plastic surgery. It can also be used to compare surgical and medical interventions, which is something we cannot do with the traditional physiologic measures (i.e., improvement of grip strength in hand surgery compared with reduction of blood, urea, nitrogen level of 10 points in a nephrology patient). The quality-adjusted life-year is perfectly suited when quality of life is the most important outcome and the intervention affects morbidity and mortality. It can also be used when only morbidity is involved.

Quality-adjusted life-years can potentially be used to inform broad resource allocation decisions among groups in the population or the choice of treatment for individual patients or patient groups. For resource allocation decisions, the quality-adjusted life-year can be used in two different types of economic evaluations. One is the prospective economic evaluation in which we collect primary data (costs and quality-adjusted life-years simultaneously with means and standard deviations).\textsuperscript{17} The second type of economic evaluation is retro-
In this type, we pool the evidence from the literature and execute a probabilistic analysis called decision analytic modeling. Decision analysis involves complex modeling such as one-way or two-way sensitivity analysis and Markov modeling. Complex Markov modeling is best used with recurrent illness, such as chronic osteomyelitis of the metacarpal bone or recurrent lymphangitis of chronic lymphedema. Performing this type of analysis requires training and statistical support. The results from economic evaluations are expressed as cost per quality-adjusted life-year gained, allowing us to compare quite disparate procedures whose outcomes are completely different from the ones we encounter in plastic surgery. When a cost-utility analysis is conducted parallel to a methodologically sound randomized controlled trial, it provides the strongest evidence to guide the decision-making process for whether to accept or reject novel technologies.

The measurement of quality-adjusted life-years is a complex endeavor, and certainly a cookbook approach is not appropriate given the complexity of many plastic surgery domains and interventions. It is important for unseasoned researchers to solicit the help of a health economist and biostatistician at the inception of a clinical research study.

We are cognizant that the methodologic concepts mentioned in this article are new to many plastic surgeons. We made an effort to minimize the mathematical calculations to the bare minimum and explain the concept of quality-adjusted life-year with diagrams. For those who may wish to explore this subject further, we recommend Thoma et al., Drummond et al., and Gold et al.

We strongly recommend that, in addition to reporting physiologic outcomes or patient-centered outcomes based on some condition-specific quality-of-life scale, investigators start measuring utilities from which we can calculate quality-adjusted life-year gains from our surgical interventions.


