Quality-Adjusted Life-Years, Comparative Effectiveness in Cancer Care

By Kerry L. Kilbridge, MD, MSc | 13 mai 2010

ABSTRACT: Recently, the American Recovery and Reinvestment Act (ARRA) set aside $1.1 billion for comparative effectiveness research (CER) to investigate what healthcare strategies and interventions offer the greatest benefits to individual patients and the population as a whole. The Institute of Medicine has identified CER in cancer care as a high priority research focus for ARRA funding. The ability to measure quality of life will be central to CER in oncology because survival and disease-free survival do not adequately capture outcomes important to policy makers, physicians, and patients. There are two ways to measure quality of life: descriptive health status and patient preference weights (utilities). However, only patient preference weights can be incorporated into the economic analysis of medical resources and be used in the calculation of quality-adjusted life-years (QALYs). Some of the advantages and limitations inherent in measuring quality of life with descriptive health status and patient preference weights are discussed. Both types of measurements face health literacy barriers to their application in underserved populations, an important concern for CER in all medical fields.

In February 2009, President Obama signed into legislation the American Recovery and Reinvestment Act (ARRA) in which $1.1 billion was set aside to support comparative effectiveness research (CER). Funding for CER will be awarded through the Office of the Secretary of Health and Human Services (HHS), the Agency for Healthcare Research and Quality (AHRQ), and the National Institutes of Health. Almost immediately, the Friends of Cancer Research (FCR)—a nonprofit organization promoting research collaboration and public-private partnerships—assembled a committee of senior academic investigators, clinicians, and advocates to provide commentary on CER. The committee published a set of recommendations based on the needs and experiences of the oncology community: Improving Medical Decisions Through Comparative Effectiveness Research: Cancer as a Case Study.

In its statement, the Friends of Cancer Research encourages broad support for the expansion of CER and the early application of research findings to clinical cancer care.[1] These recommendations have now been endorsed by more than 25 medical groups. A few months later, at the behest of the United States Congress, the Institute of Medicine (IOM) convened a committee to help define a research agenda for investigations to be supported through ARRA funds. The recommendations of the IOM committee are
contained in its report, Initial National Priorities for Comparative Effectiveness Research, which was released in June 2009 at the same time a second national report was released from the Federal Coordinating Council for Comparative Effectiveness Research.[2,3]

The IOM committee provided a working definition of CER by reconciling the definitions used by the Congressional Budget Office, prior studies by the IOM, the American College of Physicians, and the Medicare Payment Advisory Commission: “…the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health-care at both the individual and population levels.”[2]

In other words, the primary goal of CER is to establish what treatments and healthcare strategies work, for whom, and how best to implement them to improve national health, as well as the health of individual patients.

The fact that cancer is common, results in widespread morbidity and mortality, and that cancer care is very costly and rapidly evolving, assured its prominence in the IOM lists of priorities. Indeed, 7 of the 25 first-tier projects involve cancer research.[2]

The Scope and Objectives of Comparative Effectiveness Research

Controversy abounds over what constitutes CER, but some areas of consensus are starting to emerge. First, both the IOM and the Federal Coordinating Council reports take pains to specify that CER does not require cost data or cost-effectiveness analysis and that CER is not used for access or coverage decisions.[2,3] However, the reports do not exclude pertinent economic considerations; the IOM states that “the overall value of a strategy can be understood best by considering costs and benefits together…When CER examines differences in costs as well as outcomes, its aim is to identify the approach that offers the better value; it does not necessarily promote or favor low-cost care.”[2]

Second, it is widely agreed that studies in CER need to focus on rigorous, head-to-head comparisons to identify best possible health-care outcomes. These comparisons should not be limited to exploring whether one drug or another is more efficacious in a phase III clinical trial setting. For CER to fulfill its promise of improving individual and population health, its research agenda must cast a wider net that includes comparing active interventions and systems of healthcare in addition to assembling healthcare strategies from a synthesis of different data sources.[1-4]

Third, after a treatment or process is shown to be effective, CER should include studies aimed at the more rapid dissemination and implementation of study results, as well as best practices that are already shown to be effective, but are not widely applied.[1-3]

Fourth, CER should emphasize effectiveness rather than efficacy. Research needs to address real-world settings and practical delivery of healthcare among a more representative cross-section of patients than the insured, more educated, younger patients that typify clinical trial enrollees in an academic cancer center.[1-3] Similarly, CER is not intended for drug development trials that are designed to establish the required safety and efficacy metrics for Food and Drug Administration approval. In order to demonstrate efficacy, such trials frequently restrict the study population to a homogenous group without comorbid illness that bears little resemblance to real-world patients, and therefore the results cannot be generalized to community clinics.[1-3]
Fifth, it is freely acknowledged that new research methodologies will be needed to increase the relevance and broaden applicability of CER for the average patient. Some of the new methodologies under discussion include pragmatic clinical trials in community-based practice settings, cluster-randomized interventions, and more nimble Bayesian adaptive designs for clinical trials that allow updating trial design in response to information generated during the investigation.[1-5]

The concept of pragmatic clinical trials with relaxed eligibility criteria, less stringent criteria for adherence, and easily measured outcomes that do not entail central adjudication, is particularly appealing for CER in cancer care at a community level. In these settings, away from tertiary care academic cancer centers, most clinics lack an electronic medical record and the highly developed research infrastructure that helps sustain the expense and administrative burden of fielding clinical trials.[1,4]

Sixth, a federal organizational infrastructure must be developed to coordinate prioritizing research topics across agencies, support the expansion of registries and databases, foster new methodologies, and educate the next generation of investigators in the methods utilized in CER studies.[2,3]

The Need to Measure Quality of Life in Cancer Care

Undertaking any type of CER in oncology, calls for an outcomes measurement that can capture quality-of-life (QOL) measures. Most treatment decisions in cancer care are made with palliative intent or with probabilities of cure that must be weighed against the toxicity side-effects of the treatment. Although great strides have been made in the use of targeted therapies with more favorable side-effect profiles, oncologists remain acutely aware of the clinical tradeoffs between length of life and quality of life. By its nature, comparing the benefits and harms of two or more cancer treatments long ago pushed the field of oncology beyond the metrics of mortality and disease-free survival, and toward considerations of QOL. Measuring QOL and not just survival allows patients, physicians, and policy makers: (1) to choose between treatment options without curative intent by judging whether the tradeoffs in side effects may be worth any gains in life expectancy or mitigation of cancer symptoms in the absence of cure, and (2) to choose between two treatments with similar survival benefits but different constellations of side effects.

Two Types of Quality-of-Life Measures

It is critical at this juncture to distinguish between the two different types of QOL measures encountered in the field of oncology: descriptive health status vs patient preference weights. In the medical literature, both measures are frequently referred to as QOL or health-related QOL, leading to considerable confusion. However, each type of measure has a different theoretical origin in addition to different applications and uses.

Descriptive Health Status

Descriptive health status measures are probably the more accessible and transparent of the two rubrics. Examples of some of the instruments found in cancer care include the various Functional Assessment of Cancer Therapy (FACT) scales, QLQ-C30, and the Expanded Prostate Cancer Index Composite (EPIC).[6-8] These measures look beyond physical function to evaluate quality of life in other areas that contribute to overall health. By considering questions about emotional health, social support, the ability to engage in usual activities, and limitations due to pain, descriptive health status helps to appraise QOL domains besides the physical function captured in performance status or survival assessments.
For instance, EPIC, a prostate cancer-specific health status measure, evaluates physical health but also assesses how patients perform relative to other domains affected by prostate cancer, such as urinary, bowel, and sexual functioning.[8] Such disease-specific domains affected by prostate cancer were found to be important to survivors but had not been included in more generic instruments like the QLQ-C30.[6] Outcome measurements based on health status assessment have their roots in psychometric theory.[9] Questions in each domain are scored as a subscale and then summed to a composite score that can be compared with other studies that have used the same measurement. Implicit in the scaling of health status is that changes of a certain magnitude in one domain are valued equivalently to the same size change in a different domain. In reality, patients may value one domain very differently from another domain. Some prostate cancer patients may regard severe bowel symptoms as worse than severe impotence, but the different values are not reflected in health status scores. A prostate cancer treatment resulting in severe bowel symptoms scores the same as a treatment resulting in severe impotence, when all other domains are held equal.

Other methodological issues develop when comparing health status scores across different diseases or for the same disease assessed by different measurement scales. For example, how does a FACT-B score in a breast cancer trial compare with an EPIC score in a prostate cancer trial, or to a Seattle Angina Questionnaire score from a cardiology trial?[7,8,10] Although generic health status measures can overcome some of these problems, the concern is that they may be insensitive to clinically relevant changes in disease-specific domains that are not surveyed explicitly. It can be difficult to inform either patients or policy makers when studies report QOL outcomes over different time frames using different health status scales containing different domains.[11,12]

**Patient Preference Weights**

Because the primary goal of CER is to compare results across a broad array of healthcare strategies and medical interventions to determine which offer the greatest benefits, the outcomes measure utilized in CER requires a common denominator that incorporates QOL over a standard time interval. The assessment of quality of life using patient preference weights allows both QOL (morbidity) and length of life (mortality) to be captured in a single unit of measure, known as a quality-adjusted life-year (QALY). QALYs can be used as the common denominator to compare results from healthcare interventions regardless of disease type, time frame, or medical discipline. Unlike descriptive health status measures, outcomes measured in QALYs in a 1-year breast cancer program can be compared with outcomes measured in QALYs for a 3-year prostate cancer trial, which in turn can be compared to QALY results in a cardiology intervention.[12-14]

Perhaps the greatest advantage to employing QALYs in CER, however, is that descriptive health status measures cannot be used for QOL adjustments in analyses of resource utilization. Only preference-weighted measures can be used to reflect quality of life in the economic analysis of resource use to determine effectiveness or cost-effectiveness. The theoretical origins of QALY measurement can be found in game theory and the axioms of expected utility theory in the field of economics.[15,16] The application of QALYs to medicine and healthcare was popularized in the 1970s and has grown steadily to the point that QALYs are now a well-accepted measure of health outcome.[17-19] Although it would be appealing to substitute descriptive health status measures for preference weights in an economic analysis, the two QOL metrics are not equivalent.[20,21] and doing so would violate the theoretical underpinnings of medical resource comparisons.[18,22-24]

**Preference Weights in the Calculation of QALYs**
How do we use patient preference weights to measure QALYs? As an example, let us compare two palliative treatment regimens for an advanced-stage cancer (see Figure 1). Treatment A involves a 6-month course of monthly multiagent chemotherapy that is more toxic than the alternative Treatment B, a 6-month course of monthly single-agent chemotherapy. On average, Treatment A results in a 3-month survival benefit compared with Treatment B. Regardless of which treatment they receive, after the side-effects of chemotherapy subside, patients feel better until their death from cancer some months later. The average patient survives for 9 months after completing Treatment A and 6 months after completing Treatment B. The patient’s experience of Treatment A can be divided into two different health states. The first health state is 6 months of monthly multiagent chemotherapy with relatively greater toxicity (ToxA) and decreased QOL. The second health state is 9 months free of treatment side effects with an improved QOL compared with chemotherapy, but with eventual death from progressive cancer (ProgCa A).

Similarly, the patient’s experience on Treatment B can be divided into two different health states. The first health state is 6 months of monthly single agent chemotherapy with relatively less toxicity (ToxB), followed by the second health state of 6 months free of treatment toxicity with an eventual death from progressive cancer (ProgCa B). The health state for the time period following Treatment B can be described identically to the health state following Treatment A; however, the ProgCa B health state is 6 months in duration, 3 months shorter than the ProgCa A health state. The benefit in QALYs for each treatment can be calculated by assigning each health state a patient preference weight on a scale of 0 (painless death) to 1 (perfect health). These weights are also known as utilities. A preference weight of 0.5 assigned to a health state means that time in that state is valued only half as much as time in perfect health. By multiplying the duration of each health state in years by its assigned preference weight and summing the products over the entire duration of an intervention, quality-adjusted survival measured in QALYs can be established for any given treatment strategy.

If a treatment course is graphed by time on the X-axis and quality of life on the Y-axis (measured by preference weight or utility), then quality-adjusted survival for a treatment is equivalent to the area under the curve (see Figure 1). How patients value the more severe toxicity of Treatment A compared with the milder toxicity of Treatment B will determine whether Treatment A’s 3-month survival benefit is worth the tradeoff in increased toxicity. A full decision analysis for more complicated strategies involves defining a set of exhaustive, mutually exclusive health states for all possible outcomes, multiplying the preference weights for each health state by the probability that these outcomes will occur, and multiplying that product by the number of years spent in each health state. These results are summed across each strategy to obtain an overall value in QALYs. The strategy resulting in the greatest number of QALYs would be expected to yield the greatest benefit in terms of QOL.[22-24]

Three Different Methodologies to Elicit Preference Weights

There are three methodologies for obtaining preference weights for a particular health state: (1) the standard gamble, (2) the time trade-off, and (3) the visual analog scale (VAS). The standard gamble is felt to adhere most closely to the theoretical underpinnings of expected utility theory, but many researchers believe that it is the most cognitively demanding of the three.
To determine a patient’s preference weight for the ToxA health state using the standard gamble, patients are asked to imagine living with the toxicity of Treatment A for the rest of their lives and to think about what they would be willing to risk to regain perfect health. Patients are then asked to choose between the certainty of living with treatment side-effects for the rest of their lives or risking a gamble. The gamble has some chance of the best possible outcome—eliminating the side effects of Treatment A and regaining perfect health, and some chance of the worst possible outcome—instant painless death. The chance of death (X%) and the chance of regaining perfect health (1X%) are varied iteratively until patients are indifferent between the gamble and living with treatment side-effects for the rest of their lives. At the point of indifference, the (1X%) chance of best possible outcome in the gamble is equivalent to the subject’s preference weight (utility) for the ToxA health state.

To assess preference weights using the time trade-off technique, patients are asked how much time in the health state of interest is equivalent to a lesser period of time in perfect health. Using our previous example, patients are asked to choose between two options:

**Option 1:** Living your remaining life expectancy with the side effects of Treatment A; or

**Option 2:** Living a shorter duration (Y years) in perfect health.

If a patient’s remaining life expectancy is 15 years, the point at which he or she is indifferent between 15 years with the side-effects of Treatment A or Y years in perfect health determines his preference weight (utility) for the ToxA health state, calculated as Y ÷ 15.

The VAS is the simplest and most straightforward measure of preference weights, but it is not considered a true utility in a strictly methodological sense.[11,12,25] Patients are given a 0-to-100 vertical or horizontal scale spanning 100 millimeters and told that 0 represents death and 100 represents perfect health. Using the example of the toxicity of Treatment A, patients are asked to choose the position Z between 0 and 100 that shows how much better ToxA is than death and how much worse ToxA is than perfect health. The preference weight is the distance from 0 to Z, divided by 100.[12,22-24,26,27]

Preference weights obtained by the different methodologies are not equivalent. A pattern emerges in which VAS weights tend to be lower than time trade-off weights for the same health state, which in turn, tend to be lower than the weights obtained through the standard gamble process. This phenomenon has been attributed to the fact that rating a health state using the VAS does not involve risk. There is little disincentive to choosing an 80 on the VAS, equivalent to a preference weight of 0.80, whereas a preference weight of 0.80 elicited through the standard gamble technique means the rater is willing to risk a 20% chance of death with an 80% chance of regaining perfect health. It is thought that time trade-off preference weights are intermediate between VAS and standard gamble weights because the time period traded for perfect health is at the end of the rater’s life expectancy at a future date, without an immediate risk of death as in the standard gamble.[12,18,23,26]

**Which Preference Weights to Use in CER**

Which preference weights should be used in a comparative effectiveness analysis? The answer depends on where the results of the analysis will be applied and whose resources are involved. If be applied at a policy level with societal resources at stake, then the recommendation is to use the mean societal preferences from members of the general public in the analysis. If results will be used to optimize decision-making for a certain population, colon cancer patients for instance, then the recommendation is
to use the mean preference weights elicited from a representative group of colon cancer patients. If results will be used in individual decision-making, then the preference weights of that particular individual should be used rather than the mean preference weights of a population.

The distinction is made for two reasons. First, there are often wide standard deviations in preference weights that discourage substituting one individual’s preferences for another individual. Second, patients as a group tend to value health states associated with their disease higher than members of the general public. As a result, the QOL benefit accruing to an intervention that decreases cancer-related outcomes will appear smaller when the preference weights of cancer patients are used than when weights obtained from the general public are used.

This difference in values has been attributed to the observation that patients of any type tend to accommodate to their decreased QOL due to illness. Consequently, a cancer patient will value his or her own health, living with cancer, higher than members of the general public who have never experienced that health state.[12,18,23,28]

Limitations in the Measurement of Preference Weights and QALYs

There are limitations to assessing preference weights and QALYs. Although the VAS can be self-administered, both the standard gamble and time trade-off methodologies require trained interviewers to carry out the elicitations. In addition, the two latter techniques are time consuming and cognitively demanding. For both of these reasons, it becomes impractical to evaluate more than seven health states in a sitting and it is difficult to obtain preference weights in a clinical setting for decision making in real time.

The latter issue can be overcome with health state classification systems like the EuroQol or the Health Utility Index (HUI) that are composed of two components: (1) a descriptive system outlining different health domains, each domain with levels of intensity from mild to severe, and (2) a preference scoring system.[29,30]

The preference scoring system employs multiattribute utility theory to attach premeasured preference weights (using the standard gamble or time trade-off methods) from prior interviews in the general public to each level of function within the descriptive system. Using the two components together allows patients to describe their health states using just the descriptive system in a clinical setting, a relatively easy task.

The scoring algorithm then allows calculation of a global weighted preference value (utility) for that health state description based on the multiattribute utility formula derived from the extensive prior interviews in the general public. One downside to most of these classification systems is that they are generic in their descriptive systems and may be insensitive to clinically meaningful changes in disease-specific domains important to patients.[29,30] An exception is the Patient-Oriented Prostate Utility Scale (PORPUS) which can measure changes in sexual, urinary, and bowel domains among prostate cancer survivors.[31]

Of greatest concern, however, is that the levels of math skills and health literacy needed in current methods of preference assessment may diminish their accessibility in the underserved populations most affected by health disparities. Existing platforms used for the standard gamble rely on the ability to understand percentages and probabilities.[27,32-34] All three techniques assume the ability to read health state descriptions that frequently include medical terms.[12,23,26,27,29,35] These requirements are juxtaposed with recently published findings in a study population of 105 underserved African American men attending low-income clinics in rural Virginia.
TABLE 1

Use of Terms in Quality of Life Measures & Percent Comprehension Among Underserved Patients

Research interviews exploring the comprehension of common prostate cancer terms documented an average reading level of fourth to sixth grade, with 27% of patients reading at a third grade level or lower. The medical terms used to characterize the side effects of prostate cancer treatment were poorly understood (see column 7 in Table 1). Less than 50% of patients understood the terms “erection” or “impotent.” Only 5% of patients understood the term “incontinence,” and just 25% understood the term “bowel habits.” Just 20% of men interviewed could calculate a simple fraction (two-thirds of nine) and a percentage (30% of 10 and 30% of 100). Many of the men had difficulty locating important structures on a simple anatomic diagram, such as the prostate, bladder, and rectum.[36]

These findings suggest that there are enormous barriers to measuring QOL outcomes in a group of patients that suffers a two- to three-fold increase in prostate cancer mortality compared with their white counterparts.[37] Moreover, there are similar barriers to measuring descriptive health status as well. The most widely used prostate-specific measures are self-administered and require sixth to ninth grade reading skills. Even if they are read aloud, their health descriptions include many poorly understood medical terms describing urinary, bowel, and sexual function (see Table 1).[8,31,38-40]

Given that the IOM has identified “Comparing the effectiveness of interventions to reduce health disparities in … cancer…” as a first-tier priority in the national agenda for CER, it is imperative that we make certain we can accurately measure QOL outcomes among underserved populations.[2]

Conclusions

National priorities for CER include important questions in the optimal delivery of cancer care. Quality-of-life considerations will be central to the investigations that address these questions. Both descriptive health status and patient preference weights are valuable quality-of-life metrics, but the latter can be used to measure QALYs in comparative economic analyses. Among the limitations of both types of QOL measures are the serious health-literacy obstacles that impede their application in some of the underserved populations that are most affected by health disparities.

References


http://www.cancernetwork.com/practice/content/article/10165/1568497


