QUALITY OF LIFE ASSESSMENT AND RESEARCH ON HEALTH CARE OUTCOMES

Joan Stelmack OD, MPH, FAAO

Chapter Overview

Based upon work by Donabedian, the evaluation of the quality of health care includes three components: structure, process and outcome. (1) Structure refers to the equipment, facilities and organization of health care. Process refers to the treatment components and the content of the interaction between patient and provider. Outcomes refer to the changes in patients’ health status after treatment. To evaluate outcomes, data analysis is performed to assess the risks and benefits of diagnostic and treatment procedures and to identify those procedures that are the most successful. There are many measures of health outcomes e.g. mortality, morbidity, patient satisfaction, disability and quality of life. Quality of life is increasingly used as an outcome measure in both clinical trials and observational studies. This chapter is focused on the evaluation of health outcomes using questionnaires that measure health-related quality of life. A case study describing the Veterans Affairs Low Vision Intervention Trial demonstrates the use of outcomes measures to evaluate the effectiveness of a new low vision outpatient program.

Objectives

On completion of this chapter, the reader should be able to:
1. Explain efficacy and effectiveness.
2. Define health-related quality of life.
3. Explain the different types of health-related quality of life measures.
4. Understand what makes a good health-related quality of life questionnaire.

Public Health Principles: Assessing Health-related Quality of Life and Health Outcomes

“Whereas epidemiology usually measures the diseases that occur from exposures to toxic agents and chemicals, outcomes research focuses on health effects of exposure to medical interventions.” (2) The terms efficacy and effectiveness are frequently used to describe treatment outcomes. (3) Efficacy refers to whether or not a treatment works in controlled clinical conditions. Effectiveness refers to whether or not a treatment works in real life. A new glaucoma medication that is administered three times per day is a good example. The glaucoma medication may be efficacious for patients in a nursing home where medication is administered by nurses; it may not be effective outside the nursing home for frail geriatric patients who have difficulty inserting the eye drops independently. A treatment must be efficacious before it can be determined to be...
effective. A treatment that does not work in an ideal setting can not be expected to work in a setting that is less than ideal.

**Models of Health Care**

Health care has recently evolved from a “disease-based” to a “patient-centered” model (4-AHRQ). Previously, physicians used medical tests results and their clinical experience make most treatment decisions. However, the test-results physicians find useful may correlate poorly with the measures of functional capacity and well being that are important to patients. As an example, a patient may have 20/20 visual acuity with a low vision device, but find the slower reading speed and reduced field of vision inadequate to meet his daily reading needs. A second limitation is that physicians do not always agree on the most appropriate treatment for a condition and outcomes may differ based upon the treatment that is prescribed. Patients with the same clinical diagnosis will value treatment options and outcomes differently based upon their individual needs. As an example, a patient with an early cataract who is employed as a truck driver may place a higher value on a small improvement in visual acuity after cataract surgery than a patient with an early cataract who does not report any difficulty driving.

In the “patient centered” model of care defined by the Institute of Medicine, “care is respectful of and responsive to individual patient preference, needs, and values and ensuring that patient values guide all clinical decisions.” (5) Patients are encouraged to consult with medical professionals to obtain information on available treatments as well as likely outcomes associated with each treatment option and to use this information to make “informed health care decisions.” Patients also identify the most important outcomes of health care and determine whether or not they have been achieved. In addition to clinical measures, treatment outcomes should include how well people function, their experiences with care and changes in quality of life. (4) Questionnaires are frequently used to measure treatment outcomes by capturing patients’ perception of their health status and quality of life before and after treatment.

**Definitions of Quality of Life and Health-Related Quality of Life**

Quality of life is difficult to define. (6-8) Different people may place a higher value on aspects of life and quality of life may mean different things to different people. The World Health Organization Quality of Life Group’s definition is “… an individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, and standards and concerns.” (9) Quality of life is “a broad-ranging concept affected in a complex way by the person’s physical health, psychological state, level of independence, social relationships and their relations to salient features of their environment.” (9) Some important aspects of quality of life, e.g. economic, political, cultural, aesthetic or spiritual that influence health, are not under the
purview of physicians and health care systems. Health-related quality of life is the term used to describe quality of life related to health care. (6)

According the National Center for Disease Prevention and Health Promotion, the concept of health-related quality of life is used in public health and medicine to “refer to a person or group’s perceived physical and mental health over time.” (10) Public health professionals track health related quality of life to measure the effects of disorders, disability and disease in different populations whereas physicians are more likely to measure the effects of illness on everyday life for individual patients. (10)

**Questionnaires**

Questionnaires can be used to gather large amounts of data quickly. (6) Questionnaires are also called as *instruments*. They include single or multiple questions that are referred to as *items*. Patients respond to each item with a *dichotomous answer* (e.g. yes or no or true or false) or with a *polytomous rating* (e.g. choosing a response from a list of ordered response categories such as ratings of difficulty or importance). Related items that assess the same variable are often grouped into *domains, dimensions or subscales*. Questionnaires may measure a *single dimension or multiple dimensions* of HRQoL. Dimensions that are frequently measured include physical (disease symptoms and treatment), functional (self-care, mobility, activity level, and activities of daily living), psychological (cognitive function, emotional status, well-being, life satisfaction and happiness) and social (social contact and inter-personal relations).

Researchers often use a *test battery* or a *scale*. A *test battery* is comprised of several questionnaires, questions, rating or items and the responses that are not summed or weighted. A *scale* is a series of self-report questions, ratings or items that are used to measure a concept. The response categories for the items are all in the same format so that they can be summed or weighted.

**WHAT MAKES A GOOD HEALTH RELATED QUALITY OF LIFE INSTRUMENT?**

Unlike the physical measurements, e.g height and weight, quality of life can not be directly observed and measured. It is a latent variable that is inferred from other variables that can be measured. (11) Questionnaires that measure quality of life must be held to the same standards as physical measurements. (11) The essential features of these measurements are unidimensionality, hierarchal order and equal scaling. (12) A ruler is frequently used to explain this concept. (13) The ruler is unidimensional. It measures only one dimension, length. When measuring objects, we think of a hierarchal order of dimensions from shorter to longer length. The ruler serves as a standard with equal scaling that does not change whether we are measuring the length of a block of wood or a bolt of fabric.

The Early Treatment of Diabetic Retinopathy Study (EDTRS) Visual Acuity Chart is frequently used by optometrists to measure distance visual acuity. We can compare measurements made with this chart to measurements made with a questionnaire to assess visual function. (14) Letters on the EDTRS chart are
larger at the top of the chart and easier for most patients to read. The letters become progressively smaller and harder to read toward the bottom of the chart. If the acuity chart has a large range of letters with small steps in letter size between each line, visual acuity can be measured with precision for patients with a wide range of visual acuity. Because a vision function questionnaire measures our ability to perform tasks, it is important the scale be calibrated to have a range of tasks from easy to hard (e.g. reading large print headlines is easy for most patients while reading small print is harder) to measure patients with a wide range of visual function. In order to make precise measurements, the intervals between task difficulties must be small.

The first step in development of a questionnaire to assess changes in health-related quality of life following medical treatments is to identify the relevant quality of life issues and the intended patient population. (15-16) Questions or items selected must be important to the patient with the illness or condition and relevant to the sample of patients who are being studied. As an example, the questions selected for patients with cancer who are undergoing chemotherapy may not be relevant to patients with macular degeneration who are being treated with anti-VEGF therapies. Relevant quality of life issues are identified through literature reviews, in-depth interviews of subject matter experts and patients with the disease, or focus groups. The development of item content for the NEI VFQ is a good example. Mangione et al. conducted 25 focus groups at 5 U.S. medical centers to determine the impact of vision problems from cataract, age related macular degeneration, diabetic retinopathy, cytomegalovirus retinitis, and glaucoma on daily living. (17) Although some problems reported were disease-specific, problems in reading, driving, general problems with seeing clear and mental health complaints caused by poor vision were reported across conditions. Ease of performance, psychological stress and complete inability to perform an activity were frequently reported. The researchers developed a comprehensive list of questions in these areas that might be included.

Short versions of the questionnaire with different wording and layouts are pre-tested on individuals from the target population. (15 -16) A dichotomous or polytomous response scale is chosen for each item or group of items. The number of items that are generated must usually be reduced to a smaller number although it is important that the scale includes enough items to address the subject matter. Items that are repetitive, confusing or not relevant to a significant percentage of patients are eliminated or rewritten to improve efficiency and reduce the length of the questionnaire.

Questionnaires are pilot tested with a diverse sample of patients from the intended population. Items that discriminate poorly, are unreliable or invalid are removed. Statistical procedures are used to reduce the number of items. The items that are retained should be useful to differentiate between people who have a better HRQoL and those who have a worse HRQoL. If the questionnaire is to be used to evaluate change in HRQoL over time, it must include items that demonstrate change.

Developers usually attempt to group items into different domains. (13) Many HRQoL questionnaires use Likert scales, classically 5 point scales on
which respondents signify their level of agreement with a statement e.g. strongly agree, somewhat agree, neutral somewhat disagree and strongly agree. A numerical rating called a raw score is given based upon the rank of the response. The numerical ratings for the ranks are treated as if they are quantities and are added together. Summary scores are developed for each subscale by summing or averaging item scores. Traditional methods of adding up the raw scores to form a total score may not be valid as we do not know the value that each rank represents. Taken at face value these numbers have no meaning.

Once a questionnaire is developed, its’ performance must be tested. (15-16) Questionnaires that are designed to measure health-related of life must be valid and reliable. Construct validity is the extent to which the range of items in the questionnaire is sufficient to address the element of quality of life that is being tested in the population of subjects that are included. Criterion validity or accuracy refers to the ability of the instrument to measure what it proposes to measure. This is usually established by comparing the new instrument to a gold standard. As there is no gold standard for quality of life instruments, validity is often determined by comparing changes in the questionnaire to changes in other clinical measures such as changes is visual acuity or visual field. (18,19)

Reliability, also referred to as reproducibility, refers to measurement of the same construct over different administrations of the instrument (15-16). Reliability is measured by administering the instrument to a group of subjects who are stable and to a group who have experienced changes in their condition. It is important to confirm that the scores on the questionnaire change when the condition changes and that the scores are consistent when the condition is stable. The usefulness of the instrument to measure change must be established prior to conducting a clinical study. If there is no change after treatment, the researchers will not know whether the treatment was not beneficial or the instrument was not able to demonstrate change. An example, the Veterans Affairs Low Vision Visual Functioning Questionnaire (VA LV VFQ-48) was developed by Stelmack et al. to use in measuring outcomes of low vision rehabilitation in clinical trials. (20) The VA LV VFQ-48 was administered before and three months after low vision rehabilitation in a control group that did not receive low vision treatment, programs that prescribe and dispense low vision devices with minimal training in their use, and a comprehensive inpatient blind rehabilitation program. Results indicated that the control group did not show improvement and the outcomes measured were smaller in the programs offering less services when their outcomes were compared to the comprehensive blind rehabilitation program. The research team concluded that the VA LV VFQ-48 was sensitive enough to be used in clinical trials to compare different service delivery models.

**Approaches to Measurement of Health Related Quality of Life**

HRQoL measures are classified as *generic* or *specific*. (6, 21) Specific measures focus on important aspects of quality of life that are relevant to the patients being
studied whereas generic measures address different populations and cover many health issues.

**GENERIC MEASURES**

There are three classifications of generic measures *single indicators, health profiles* and utility measures (6,21). A *single indicator* includes one question e.g. how are you feeling today? The advantage of single indicators is that they can be administered quickly with minimal patient burden. The disadvantage is that one question may not be sufficient to measure a concept. An example of a single indicator is CDC’s four “Healthy Days” Measures that are used to measure the health of adults in the community (22). Each measure is comprised of a single question. The first question asks the respondent to rate their health as *excellent, very good, good, fair* or *poor*. The second measure asks for an estimate of the number of dates during the past 30 days that physical health was not good. The third measure asks for an estimate of the number of days during the last 30 days that mental health was not good. The fourth measure asks for an estimate of the number of days during the past 30 days that poor physical or mental health prevented performance of usual activities such as self-care, work or recreation. Summary indices of healthy and unhealthy days are calculated. The “Healthy Days” measures are used on the Behavior Risk Factor Surveillance System BRFSS and the examination component of the National Health and Nutrition Examination Survey (NHANES) to identify health disparities and track population trends and at the state and local level.

*Health profiles* combine scores from multiple dimensions of quality of life into a single number (6,21). The major advantage of health profiles is that they can be used in any population to compare the relative impact of different health care interventions. The major limitation is that they may not be sensitive to changes in specific disease conditions. The Short Form-36 (SF-36) is the most widely used generic health profile. (23-24) It is referenced in over 4,000 publications and in articles that describe over 200 conditions and diseases. Most of the items selected for the SF-36 were used in other questionnaires. It is common practice for researchers to include questions that have been used in other questionnaires because the questions have already been tested repeatedly and are psychometrically sound. The 36 items on the SF-36 are divided into eight scales that profile functional health and well being. Summary scores are also provided for physical and mental health. The physical health measures include the *Physical Functioning, Role-Physical, Bodily Pain* and *General Health* scales. The mental health measures include the *Vitality, Social Functioning, Role-Emotional* and *Mental Health* scales. Extensive psychometric evaluation of the instrument has been performed and reported in the literature. General population norms are also available for comparison purposes. A shorter version, the SF-12, was also constructed from a subset of 12 SF-36 items. Short form outcome measures are frequently used to reduce patient burden and costs associated with administration of questionnaires.
Utility assessment is another generic measure of HRQoL. (6,25-26) Utilities measure patients’ preference for different health states. Rating scales, time trade-off, standard gamble and pre-scored instruments are the methods most frequently used to measure utilities. Utility scores range from 1.0 for perfect health to 0 for death. Patients use rating scales to rank health states associated with different health outcomes from most to least preferred and place them on a scale such that the sizes of the spaces between the health states reflect the patient’s preferences. In the time tradeoff, respondents are asked to consider a series of trade-offs to determine their preference for a particular health state. As an example, a patient with diabetes could be told that based upon his age he has a life expectancy of 20 years. He would then be asked if he would be willing to give up a certain number of years of his life to be restored to perfect health. If the answer is no, the question is repeated and the number of years of life to be sacrificed reduced. If he responds yes, the time is lengthened. The end point is reached when the patient expresses no preference between a shorter lifespan with perfect health vs living with his diabetes over his life expectancy. In the standard gamble, the individual is offered treatment with two possible outcomes. Either the patient is returned to perfect health with life for an additional number of years (probability p) or the patient dies immediately. In either case the probability of each scenario is varied until the individual can not choose between the alternatives.

The quality adjusted life year (QALY) is a measure includes the years of life lived with the burden of disease and the quality of life lived (25,26). QALYs are the arithmetic product of life expectancy and a measure of the quality of the remaining life years. QALYs can be combined with the costs of interventions to allow comparisons of the benefits gained from different interventions.

SPECIFIC MEASURES

Specific HRQoL measures focus on aspects of health status that are of primary interest. Included in this approach are disease specific (e.g. vision, arthritis, cancer, and diabetes), population specific (e.g. women with urinary incontinence), function specific (e.g. visual function), condition or problem specific (e.g. pain) measures. (6) These scales may be more responsive because only aspects of HRQoL relevant to the patients who are being studied are included. Specific measures are not used to compare the burdens of different diseases across populations, although generic and disease specific instruments are often included in the same study to allow comparisons to other populations and patients with other diseases.

The National Eye Institute Visual Functioning Questionnaire (NEI VFQ-25) and the Vision Function (VF-14) are examples of specific measures that are frequently used in vision research (71-18,25). The National Eye Institute contracted the RAND Corporation to develop the NEI VFQ to use in clinical trials to assess patient perception of visual function and the influence of visual disability on multiple dimensions of health-related quality of life. A 51 item NEI VFQ was constructed and later shorted to a 25 item version. The NEI VFQ-25 has been used in a wide range of populations based studies and clinical
interventions varying from surgical treatment and pharmaceutical therapies to driving habits and self-management of macular degeneration. The subscales of the NEI VFQ-25 include: general health status, difficulty with general vision activities, difficulty with activities at distance, difficulty with activities at near, role difficulty, social functioning, and dependency on others, mental health symptoms, driving difficulties, limitations on peripheral vision, color vision and ocular pain.

Visual function is frequently used as a surrogate measure for HRQoL. Some of the scales that were developed to measure visual function were initially developed for clinical trials of specific eye diseases. As an example, both the Activities of Daily Vision Scale (ADVIS) (Mangione27 and the VF-14 (Steinberg 19) were developed as indices of functional impairment in patients with cataract. The VF-14 contains a list of 14 activities including: read small print, such as labels on medicine bottles, a telephone book or food labels; reading a newspaper or book; reading a large-print book or newspaper or the numbers on a telephone; recognizing people when they are close to you; seeing steps, stairs or curbs; reading traffic, street or store signs; doing fine handwork such as sewing, knitting crocheting, or carpentry; writing checks or filling out forms; playing games such as bingo, dominos, card games, or mahjong; taking part in sports such as bowling, handball, tennis or golf; cooking; watching television; daytime driving and nighttime driving.

Administration of Questionnaires

The method chosen for administration of questionnaires often depends upon the financial resources available and the length of the instrument (28) Questionnaires are often self-administered to reduce staff time and to provide privacy for the respondent. They can be either be completed in a clinical setting or mailed. Longer questionnaires or questionnaire batteries are often administered in a personal interview in the patient’s home or by telephone. The advantage of interviewer administration is that help can be given. Questions can be read to visually impaired patients and any ambiguities in the questions can be clarified by the interviewer. The costs of interviewer administration are usually considerably higher than self-administration. Regardless of the method of administration, written instructions are essential to standardize the administration across subjects and to reduce biases from the interactions of patients and interviewers.

Application of Public Health Principles

CASE STUDY

The Hines VA Hospital is interested in designing a study to compare the effectiveness of two different low vision programs that serve legally blind veterans. Our assignment is to recommend an outcome measure for the study.
Public Health Questions for the Case Study

1. How should we measure the outcomes of the low vision programs?
2. Should we use a generic or a disease specific measure?
3. How should we administer the outcome measure?

How should we measure the outcomes of the low vision programs?

Clinical trials to evaluate new treatments of eye diseases frequently use two or three lines of improvement in visual acuity as an outcome measure. Visual acuity is not; however, an appropriate measure for low vision outcomes as low vision rehabilitation does not restore vision or improve visual acuity. Low vision services improve patients visual functioning by providing strategies and assistive devices to enhance remaining vision or substitute for vision loss.

Previous low vision outcomes studies have used many different criteria for success e.g. visual acuity with low vision devices, clinical performance measures and questionnaires for patients to self-report of their quality of life. Although some studies have measured speed and or accuracy of tasks in the clinic these clinical measures of performance may not reflect patients’ perception of their functioning. Every low vision practitioner has patients who benefit from low vision reading devices during clinical evaluations and training, but rarely, if ever, use these devices. A patient may consider the slower reading speed and amount of effort needed to visually read or the pain experienced holding the magnifier with arthritic fingers as a minimal improvement in function. Given the emphasis on patient centered outcomes, we want to measure effects that patients with low vision experience and care about such as their changes in ability to function at home and in the community after rehabilitation. Questionnaires are frequently used for patients to self-report their quality of life.

Clinical pearl: For chronic conditions where a cure is not possible, it is important to assess patients’ perception function and well being and to include questionnaires that capture patients’ self-report of their quality of life.

Which should we use: a generic or a disease specific questionnaire?

A generic outcome measure e.g. the SF-36 addresses many different health issues. These measures are generic in determining the effects of various diseases on quality of life. Their ability to measure changes in quality of life for patients with specific diseases is limited. We expect that there would be little change in the SF-36 scores if the instrument was administered before and after a low vision program. A better choice would be a specific outcome measure e.g. the NEI VFQ-25, the VA LV VFQ-48. Both of these instruments were developed to evaluate quality of life relative to eye diseases and their treatment. A literature search confirms that both instruments have been used to measure outcomes of low vision rehabilitation and that they are both sensitive to the changes that occur with low vision treatment.
Clinical pearl: Generic measures enable us to measure and compare the health status and quality of life of patients with different diseases. Specific measures are needed to measure and compare the health status and quality of life of patients with a specific disease as these measures are focused on the aspects of the disease that are most important to the patients who have the disease or condition we are studying.

How should we administer the outcome measure?

We need to consider the intended population – elderly veterans with severe vision loss and the goal of our study, to measure patients’ perception of changes in their quality of life after low vision rehabilitation. The easiest and most cost-effective method would be to mail the questionnaires to the veterans. There are several problems with this approach. Veterans who do not have low vision devices will not be able to fill out the questionnaire independently before they enroll in the low vision rehabilitation programs. Veterans who were not successful in the programs and veterans with severe vision loss may still have difficulty reading and filling out the questionnaire. If family members or other caregivers assist the veterans, we may capture family and caregivers’ impressions rather than the veterans self-report of their ability to function. It is also likely that some of the questionnaires that veterans cannot read will not be returned. Thus, the response rate is likely to be low, some items may be missing and there may be errors of misunderstanding. We could send an interviewer to the veterans’ home to administer the questionnaires. However, this approach is very expensive and not feasible given the funding that is allocated for the study. A better alternative is telephone administration of the questionnaire. The research assistant will contact the veteran by phone and make an appointment to administer the questionnaire at a time that is convenient for the veteran. We expect that there will be few if any missing items or misunderstandings and our response rate should be higher.

Clinical pearl: The strengths and weakness of different modes of administration for questionnaires should be carefully considered when planning clinical studies.

Public Health Burden Addressed

Low vision rehabilitation has the potential to improve functional ability and increase independence of elderly veterans. Programs should be compared to determine which service delivery models are most effective. The data suggests that the most appropriate outcome measure to use in comparing the effectiveness of the two low vision programs would be a specific quality of life questionnaire that is administered by telephone.
Study Questions:

1. Efficacy measures:
   A. Treatment in ideal conditions.
   B. Treatment in real-life conditions
   C. Any treatment response
   D. All of the above

2. The essential features of measurement include:
   A. Unidimensionality
   B. Hierarchical order
   C. Equal scaling
   D. All of the above

3. An example of a generic quality of life instrument is:
   A. NEI VFQ-25
   B. VF-14
   C. ADVIS
   D. SF-36

4. Which measure combines quality and quantity of life?
   A. QALY
   B. SF-36
   C. NEI VFQ-25
   D. Mortality

5. Which methods are used to measure utilities?
   A. Time tradeoff
   B. Standard gamble
   C. Rating scales
   D. All of the above

Take Home Conclusions

- Quality of care has three components, structure, process and outcomes. Outcomes refer to the changes in patients’ health status after treatment.
- Two terms, efficacy and effectiveness are used to describe treatment outcomes. Efficacy refers to whether or not a treatment works in controlled conditions whereas effectiveness refers to whether or not a treatment works in real life.
- The “patient-centered’ model of health care is responsive to individual patient preferences, needs and values.
- Quality of life related to health care is described as health-related quality of life. While there are many definitions of HRQoL, it can be generally...
described as a person or groups perceived functioning, mental and physical health over time.

- Measures of HRQoL are classified as generic or specific. Generic measures focus on important aspects of quality of life that are relevant to the patients being studied. Generic measures address different populations and cover many health issues.
- The modes of administration of HRQoL measures include: interviewer, telephone, self and surrogate responders. The strengths and weakness of the different modes should be examined when planning a study.

References


9. World Health Organization Quality of Life Group. The World Health Organization Quality of Life Assessment (WHOQOL): position paper from
the WHO Organization. Special issue on health-related quality of life: what is it and how we should measure it? Social Science and Medicine 1995;41(10):1403-1409.


