

Chapter 13

Assessment of Health Related Quality of Life

Michelle J. Naughton and Sally A. Shumaker

The major goal of randomized clinical trials is to determine the potential benefits and harms of an intervention. The benefits of most available interventions in medicine are symptom improvements. Thus, relief or reduction of symptoms is a common primary outcome in clinical trials (Chap. 3). Most of the adverse effects of interventions are also symptom-related (Chap. 12). Most changes in symptomatology are subjective and reported by trial participants, with a special form of outcomes related to various types of functioning, traditionally covered by the term health-related quality of life (HRQL) [1–4].

A person's perspectives and experiences have recently been integrated in a new term—"Patient-Reported Outcomes" [5–7], defined by the FDA as "...any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" [8].

In this chapter we focus on the traditional outcome of HRQL and discuss types of measures, their uses, methodological issues and their selection.

Fundamental Point

Assessments of the effects of interventions on participants' daily functioning and health-related quality of life are critical components of many clinical trials, especially ones that involve interventions directed to the primary or secondary prevention of chronic diseases.

M.J. Naughton, Ph.D.
The Ohio State University, Columbus, OH

S.A. Shumaker, Ph.D.
Wake Forest School of Medicine, Winston-Salem, NC

Types of HRQL Measures

Primary Measures

What is meant by quality of life varies greatly depending on the context. In some settings, it may include such components as employment status, income, housing, material possessions, environment, working conditions, or the availability of public services. The kinds of indices that reflect quality of life from a medical or health viewpoint are very different, and would include those aspects that might be influenced not only by conditions or diseases, but also by medical treatment or other types of interventions. Thus, HRQL is commonly used to mean the measurement of one's life quality from a health or medical perspective.

In general, HRQL measures are multi-dimensional to reflect different components of people's lives. Although there are some variations, there is general agreement on the primary dimensions of HRQL that are essential to most HRQL assessments [9]. These include: physical, social and psychological functioning, and participants' overall assessment of their life quality and/or perceptions of their health status.

Physical functioning refers to an individual's ability to perform daily life activities. These types of activities are often classified as either 'activities of daily living,' which include basic self-care activities, such as bathing and dressing, or 'intermediate activities of daily living,' which refer to a higher level of usual activities, such as cooking, and performing household tasks.

Social functioning is defined as a person's ability to interact with family, friends and the community. Instruments measuring social functioning may include such components as the person's participation in activities with family, friends, and in the community, and the number of individuals in his or her social network. A key aspect of social functioning is the person's ability to maintain social roles and obligations at desired levels. An illness or intervention may be perceived by people as having less of a negative impact on their daily lives if they are able to maintain role functions that are important to them, such as caring for children or grandchildren or engaging in social activities with friends. In contrast, anything that reduces one's ability to participate in desired social activities, even though it may improve clinical status, may reduce the person's general sense of social functioning.

Psychological functioning refers to the individual's emotional well-being. It has been common to assess the negative effects of an illness or intervention, such as levels of anxiety, depression, guilt and worry. However, the positive emotional states of individuals should not be neglected. Interventions may produce improvements in a person's emotional functioning, and therefore such aspects as vigor, hopefulness for the future, and resiliency are also important to assess.

Global Quality of Life represents a person's perception of his or her overall sense of well-being and quality of life. For example, participants may be asked to indicate a number between 0 (worst possible quality of life) and 10 (best possible quality of life) which indicates their overall quality of life for a defined time period (for example, in the last month).

Perceptions of health status need to be distinguished from actual health. Individuals who are ill and perceive themselves as such, may, after a period of adjustment, reset their expectations and adapt to their life situation, resulting in a positive sense of well-being. In contrast, persons in good health may be dissatisfied with their life situation, and rate their overall quality of life as poor. Participants may be asked to rate their overall health in the past month, their health compared to others their own age, or their health now compared to 1 year ago. It is interesting to note that perceived health ratings are strongly and independently associated with an increased risk of morbidity and mortality [10–12], indicating that health perceptions may be important predictors of health outcomes and HRQL, independent of clinical health status.

The dimensions of HRQL assessed in a trial should match the aims of the study. Some trials will necessitate the measurement of multiple dimensions, whereas others may suffice with the inclusion of one or two dimensions. For example, it is unlikely that in the examination of the short-term effects of hormone therapy on peri-menopausal symptoms, general physical functioning of the study participants (women in their mid-forties to early fifties) will be influenced. The inclusion of this dimension of HRQL in the trial may simply increase participant burden without benefit. It is important for investigators to indicate clearly the dimensions of HRQL used in a trial and provide a rationale for their inclusion (or exclusion), for example, deleting HRQL dimensions that might make the treatment under study “look bad”.

Additional Measures

Sleep disturbance has been related to depression and anxiety, as well as diminished levels of energy and vitality. Instruments assessing sleep habits may examine such factors as sleep patterns (e.g., ability to fall asleep at night, number of times awakened during the night, waking up too early in the morning or difficulty in waking up in the morning, number of hours slept during a typical night); and, the restorativeness of sleep.

Neuropsychological functioning refers to the cognitive abilities of a person, such as memory, executive functioning, spatial and psychomotor skills. This dimension is being more commonly assessed for a wide range of health conditions or procedures, such as the effects of a stroke, cardiac surgery, chemotherapy or multiple medications on cognitive functioning, as well as in studies of older people.

Sexual functioning measures include items regarding a person's ability to perform and/or participate in sexual activities, the types of sexual activities in which one engages, the frequency with which such activities occur, and persons' satisfaction with their sexual functioning or level of activity. These assessments are particularly important in studies in which the disease's or condition's natural history or its treatment, can influence sexual functioning (for example, antihypertensive therapy, prostate cancer surgery, or sequelae of a stroke).

Work-related impacts encompass a wide range of both paid and unpaid activities in which individuals engage. Measures of this dimension might include

paid employment (for instance, time of return to work, hours worked per week); household tasks; and volunteer or community activities. Also, among employed individuals, the impact of the inability to work or fully return to employment, as well as health and life insurance issues are being increasingly assessed.

Although the above symptoms are some of the more commonly assessed in clinical research, other symptoms may be important to measure. Again, the specific symptoms relevant for a given clinical trial will depend upon the intervention under investigation, the disease or condition being studied, the aims of the trial, and the study population [13].

Uses of HRQL Measures

For many participants, there are two primary outcomes that are important when assessing the efficacy of a particular intervention: changes in their life expectancies and the quality of their lives. HRQL measures provide a method of measuring intervention effects, as well as the effects of the untreated course of diseases/health conditions, in a manner that makes sense to both the participant and the investigator. In countries where chronic rather than acute conditions dominate the health care system, the major goals of interventions include the reduction of symptoms, and maintenance or improvement in functional status. Increasing costs of health care and prescription medications also necessitate the thorough evaluation of competing treatments for optimal health and quality of life outcomes. Thus, it is important to determine how the person's life is influenced by both the disease and its intervention, and whether the effects are better or worse than the effects of the untreated course of the underlying disease.

There are now many published studies assessing the HRQL and symptoms of participants in clinical trials. One classic clinical trial by Sugarbaker and colleagues examined 26 patients with soft tissue sarcoma to compare the impact of two treatments on physical functioning and symptoms [14]. Patients were randomized to amputation plus chemotherapy or limb-sparing surgery plus radiation therapy and chemotherapy. After all treatments had been completed and the participants' physical status had stabilized, assessments were completed to measure HRQL, economic impact, mobility, pain, sexual relationships and treatment trauma. Contrary to expectations, participants receiving amputation plus chemotherapy reported better mobility and sexual functioning than those receiving limb-sparing surgery plus irradiation and chemotherapy. Based on the results of this study, practices in limb-sparing surgery, radiation and physical therapy were modified to improve patient care and functioning.

An example of a clinical trial examining findings first noted in observational studies, which has had widespread impact on clinical care, is the Women's Health Initiative (WHI) hormone therapy trials. During the 1980s and early 1990s, observational and case-control studies suggested that the use of estrogen would decrease the incidence of cardiovascular events among post-menopausal women. In order to determine if this observation would be replicated in a large, randomized controlled

trial, the WHI was initiated in 1993 [15]. Post-menopausal women ages 50–79 at baseline were randomized to either conjugated equine estrogens plus medroxyprogesterone acetate (CEE + MPA) versus placebo if they had not had a hysterectomy, or conjugated equine estrogens (CEE-alone) versus placebo among participants who had had a hysterectomy. The trial was expected to last an average of 8.5 years. Health-related quality of life was assessed annually after trial initiation. In 2002, the trial component testing CEE + MPA was stopped early, due to higher rates of cardiovascular events and breast cancers among women in the CEE + MPA arm versus the placebo group [16]. A year and a half later, the CEE-alone component was also stopped due to adverse outcomes among women randomized to the hormone therapy group [17]. The results of these two trials have had a major impact on the care recommendations of post-menopausal women, and spurred a debate among primary care practitioners, cardiologists, and gynecologists about the validity of the WHI results [18]. One argument that was made was that although estrogen therapy may not be indicated for cardiovascular disease protection, women still reported better HRQL when taking estrogen therapy. However, the quality of life results from the WHI did not support this argument [19]. Among women randomized to CEE + MPA versus placebo, the use of active treatment was associated with a statistically significant, but small and not clinically meaningful benefit in terms of sleep disturbance, physical functioning, and bodily pain 1 year after the initiation of the study. At 3 years, however, there were no significant benefits in terms of any HRQL outcomes. Among women aged 50–54 with moderate-to-severe vasomotor symptoms at baseline, active therapy improved vasomotor symptoms and sleep quality, but had no benefit on other quality of life outcomes. Similar results were found in the CEE-alone trial of the WHI among women with hysterectomy. At both 1 and 3 years after the initiation of the trial, CEE had no significant and clinically meaningful effects on HRQL [20]. Thus, the potential harmful effects of estrogen therapy among post-menopausal women were not outweighed by any significant gains in quality of life.

More recent trials have utilized HRQL as both primary and secondary outcomes. Richardson and colleagues conducted a randomized trial to assess a collaborative care intervention versus usual care for adolescents with depression [21]. Youth between the ages of 13–17, who screened positive for depression using the Patient Health Questionnaire (PHQ) [22] on two separate occasions and met criteria for major depression, were recruited. Adolescents randomized to the intervention arm had an in-person clinic visit with subsequent regular follow-up sessions with a master's level clinician. The control group participants received their screening results and were referred to mental health services in the health care plan. The primary outcome was a change in depressive symptoms as measured by the Children's Depression Rating Scale-Revised (CDRS-R) [23] from baseline to 12 months. Secondary outcomes included the change in Columbia Impairment Score (CIS) [24], depression response (>50% decrease on the CDRS-R) and a PHQ-9 score <5, signifying depression remission. The results indicated that the adolescents in the intervention group had statistically significantly greater decreases in the CDRS-R scores than the usual care group. Both groups experienced improvement on the CIS, with no significant differences between the groups.

However, the intervention youth were more likely to achieve depression response and remission than the control group of adolescents. The results suggested that mental health treatment can be integrated into primary care services.

The Comparison of Laser, Surgery and Foam Sclerotherapy (CLASS) clinical trial examined the impact of treatment for primary varicose veins on HRQL [25]. This was a multicenter study of 11 vascular surgery departments in the United Kingdom involving 798 participants. Participants were randomized to ablation therapy, surgery, or foam sclerotherapy. For the primary outcomes, the investigators used the disease-specific Aberdeen Varicose Veins Questionnaire [26] and the generic SF-36 [27] and the Euroqol Group 5-Dimension [28] measures. Secondary outcomes were complication rates and measures of clinical success. Outcomes were assessed at baseline and 6 weeks and 6 months after treatment. The results indicated similar HRQL outcomes across the three groups, although slightly worse disease-specific quality of life was observed in the foam group as compared with the surgery group. All treatments had similar rates of clinical success, but complications were lower in the laser treatment group, and the foam group had less successful ablation of the main trunk of the saphenous vein than the surgery group. Thus, all of these examples indicate that HRQL can be used as both primary and secondary outcomes, and can have substantial impact on clinical care practices and treatment options.

Methodological Issues

The rationale and execution of a well-designed and conducted randomized clinical trial assessing HRQL is the same as for other study outcomes. The reasons for its inclusion must be specified with supporting scientific literature and the HRQL measures selected should match the specific aims and have sound psychometric properties. If HRQL measures are secondary outcomes, it is also important to have sufficient study power to detect changes in these outcomes. The double-blind design minimizes the risk of bias.

The basic principles of data collection (Chap. 11) which ensure that the data are of the highest quality are also applicable to HRQL assessments. The methods must be feasible and designed to limit missing data. Training sessions of investigators and staff should be conducted for all trials, as well as pretesting of study data collection procedures and study measures, including HRQL assessments. An ongoing monitoring or surveillance system enables prompt corrective action when errors and other problems are found.

Design Issues

Several issues must be considered when using HRQL measures in clinical trials [3, 4]. These include the characteristics of the participants, type of treatment or intervention being studied, and protocol considerations.

Study Population

It is critical to specify key population demographics that could influence the choice of HRQL measures and the mode of administration. Education level, gender, age range, literacy levels, the primary language(s) spoken, and cultural diversity should be carefully considered prior to selecting any measures. Functional limitations should also be assessed. Elderly people may have more vision or hearing problems than middle-aged persons, making accommodations to self- or interviewer-administered questionnaires necessary. Ethnically diverse groups also require measures that have been validated across several different cultures and/or languages [29]. Children generally need instruments specifically for their age group, as well as assessments from parents regarding their perceptions of their child's symptoms and physical and psychological health status.

The health status of the participant at baseline must also be taken into account in the development of the protocol and data collection procedures, including the severity of the illness, the effects of the participants' illness or health condition on daily life, symptom levels or whether symptoms are acute or chronic. Healthy or mildly ill individuals will likely be able to participate more in a trial than those with debilitating chronic health conditions. These considerations have ramifications for the burden placed on participants (and staff) in completing study requirements and data collection, or those in acute phases of an illness. Participants who are children and/or are unable to complete HRQL assessments themselves may require the use of family proxy and/or investigator or staff assessments to collect HRQL data.

It is also important to be sensitive to how the underlying condition will progress and affect the HRQL of participants in the control group, as it is to understand the effects of the study intervention on those in the intervention arm(s). The point is to select dimensions and measures of HRQL that are sufficiently sensitive to detect changes in *both* the intervention and the control group participants. Using the same instruments for both groups will ensure an unbiased and comparable assessment.

Type of Intervention

Three major intervention-related factors are relevant to the assessment of HRQL—the favorable and adverse effects of intervention, the time course of the effects, and the possible synergism of the intervention with existing medications and pre-existing health conditions. It is important to understand how a proposed intervention could affect various aspects of an individual's life in both positive and negative ways. What effects may the participant experience as a result of intervention? Some oral contraceptives, for instance, may be very effective in preventing pregnancy, while producing cyclical symptoms like bloating and breast tenderness, and in severe cases, blood clots. Dietary interventions designed to increase fruit and vegetable intake and lower dietary saturated fats may cause mild gastrointestinal

effects, which may dissipate over time. Thus, the time course of an intervention's effects is important both in terms of the selection of measures and the timing of when HRQL measures are administered to study participants. Furthermore, it is important to know the medications the participants are likely to be on prior to randomization and how these medications might interact with the trial intervention, (either a pharmacological or behavioral intervention), to influence the dimensions of HRQL.

The frequency of HRQL assessments will depend on the nature of the condition being investigated (acute versus chronic), the expected effects of the intervention, and the specific aims of the trial. Ideally, a baseline assessment should be completed prior to randomization and the initiation of the intervention. Follow-up assessments should be timed to match expected changes in functioning due to either the intervention or the condition itself. In a trial comparing a new acne skin serum with a placebo oil-free lotion for the treatment of severe acne in adolescents, assessing skin redness, sensitivity and acne reduction at only 1 and 3 weeks after baseline might not be sufficient to accurately measure the effectiveness of the intervention vs. placebo, given that severe acne may take longer than 3 weeks to show noticeable skin improvements even with known effective treatments. If the HRQL assessments are instead completed at baseline and 2 week intervals through 8 weeks, treatment effects (or non-effects) might be more accurately assessed. Thus, the timing of the HRQL assessment will affect the interpretation of the benefits (or negative effects) of the interventions.

Frequency of Assessment (Acute Versus Chronic)

In general, acute conditions resolve themselves in one of four ways: a rapid resolution without a return of the condition or symptoms; a rapid resolution with a subsequent return of the conditions after some period of relief (relapse); conversion of the acute condition to a chronic problem; or death [30]. In the case of rapid resolution, HRQL assessments would likely focus on the relative effect of the condition's symptoms on the participant's daily life. When there is a risk of relapse, a longer duration of follow-up is necessary, because relapses may have a broad impact on the participant's general functioning and well-being. If the acute problem converts to a chronic condition, the evaluation is complicated by the duration of time and the problem of how to balance participant functioning in making treatment decisions.

Interventions that have little or no adverse effects on participant function are best evaluated on the basis of their impact on survival or change in disease severity or risk. In these situations, HRQL assessments will be of lesser importance. However, when a disease or condition affects functional capacity, interventions should to be evaluated for their effects upon the participants' level of functioning and well-being. Again, in these situations, the type of HRQL instruments used and the timing of the assessments will depend on the nature of the condition, the intervention, and the expected time course of effects on the participants.

Protocol Considerations

After consideration has been given to the study population, the nature of the condition being studied and characteristics of the proposed intervention, there are additional protocol-related factors that need to be taken into account when developing HRQL collection procedures. Factors such as the venue for the proposed intervention (e.g., in clinic or hospital, community site, home, or school) and whether the intervention is done by trained staff, via computer, or using some other method, will influence the methods used to collect data. In addition, the number of participants being recruited to the trial, the number of follow-up assessment points, and the overall length of the trial (e.g., 8 weeks vs. 4 years) will have ramifications for the study design. Participants seen in clinics at regular intervals may afford easy access to completing assessments. Other modes of data collection, such as telephone, mail, or computer all have strengths and weaknesses. Telephoning participants to complete symptom or HRQL measures takes up staff time, but may involve less staffing, expense and missing data than preparing and sending a mailing to participants, tracking the responses and perhaps doing a second mailing or telephone call to increase the response rate. Interviewer administered instruments also generally provide more complete data and allow for probes and clarification. However, there may be a reluctance on the part of some participants to openly discuss some issues (for example, depression, sexuality), whereas they may be willing to respond to questions about these issues in a self-administered format. For populations with a relatively high proportion of functional illiteracy, in-person interviewer administration may be required. Interviewer administration may also be the best way to obtain information for culturally diverse populations. Interviewer-administered instruments, however, are subject to interviewer bias and require intensive interviewer training, certification, and repeat training, especially within the context of multi-site clinical trials that may be of a long duration. Thus, they can be more expensive than self-administered instruments and serious thought must be given at the planning phases of a trial regarding the trade-offs between these strategies.

On-line ascertainment has become more feasible and popular. They may not be an optimal choice, however, for those without ready access to on-line resources. Hand held devices and tablets for the tracking of symptoms are becoming more widely used, but take time to train staff and participants in their use. In addition, depending on the number of participants in a trial, obtaining these devices may be cost-prohibitive. If participants are only being assessed at 6 month intervals vs. weekly, for example, the use of mailed or on-line ascertainment may be more cost-effective.

All methods of data collections have their pluses and minuses and need to be considered in devising the most optimal methods for completing HRQL assessments economically with as little participant and staff burden as possible, while minimizing missing data. Options for data collection need to be assessed during the development of the protocol, and not as an afterthought. If HRQL assessments are

secondary outcomes, data collections procedures will need to accommodate the needs of the primary aims, but should still be approached with the same rigor and planning as the collection of primary outcomes data.

Modifying and Mediating Factors

HRQL measures may be influenced by both modifying and mediating factors. *Modifying factors* are those variables that can modify the effect of an intervention on an outcome. They can be divided into three categories: contextual, interpersonal, and intrapersonal [31]. Contextual factors include such variables as study setting or the living environment of a participant (for example, urban vs. rural, single dwelling house vs. multiunit building, clinic vs. home intervention); economic structure (e.g., national health insurance); and sociocultural variations (e.g., customs, social norms). Interpersonal factors include variables such as the social support available to individuals, stress, economic pressures, and the occurrence of major life events, such as bereavement and the loss of a job. Intrapersonal factors are associated with the individual, such as coping skills, personality traits, or physical health. *Mediating factors* are any changes, improvements or impairments to a participant's well-being that are induced by the study intervention. These are the changes that are most often assessed in trials with HRQL or symptom outcomes. For example, in a trial studying the effectiveness of aromatase inhibitors in preventing cancer recurrence among breast cancer survivors, these drugs may cause moderate to severe joint and muscle pain, which could lead to reduced HRQL and treatment adherence, although the study drug is effective in increasing overall cancer-free survival.

In addition, changes in the natural course of the disease or condition (i.e., whether the condition improves or deteriorates) must be considered in HRQL assessments, especially in trials of relatively long duration. Investigators should consider what effects the intervention or the health condition itself will have on participants' well-being, and any factors that might moderate these relationships, in order to better select and measure pertinent HRQL variables. Consideration of these factors will aid in the interpretation of study findings, and may enable investigators to explain better the results of a specific intervention.

Selection of HRQL Instruments

All HRQL outcomes must be participant-centered, and the instruments used must match the specific aims of each particular clinical trial. For example, in a study examining the impact of post-surgery swelling on physical and social activities, one would not only need to determine whether and where swelling occurs, but how much it interferes with the ability to carry out physical and social activities. Simply

measuring the occurrence and frequency of swelling, for example, would not answer the question of the effect on daily life.

Recently, there have been several reviews that have identified minimum quality standards for HRQL and other patient-reported outcomes [32, 33]. These attributes include measures with 1) a conceptual model; 2) established reliability; 3) established validity; 4) responsiveness to changes in clinical status and/or as a result of one intervention; 5) interpretability of scores; 6) cultural and language translations or adaptations; 7) feasibility in the desired setting; and 8) participant and staff/investigator burden. It is beyond the scope of this chapter to review techniques and practices used to develop HRQL measures, but references regarding scaling procedures and psychometric considerations of instruments (reliability, validity, and the responsiveness of instruments to change) may be consulted [3, 4, 34].

Types of Measures

HRQL measures can be classified as either generic (that is, instruments designed to assess outcomes in a broad range of populations), condition/disease specific (e.g., congestive heart failure, cancer) or symptom-specific (e.g., pain, anxiety) [13]. Within these categories of measures are single or multiple questionnaire items. Single questionnaire items that ask participants to rate their current severity of a symptom on a scale from 0 to 10, have the advantage of limiting participant burden and can generally be completed and understood by most people. Multiple questionnaire items provide greater information and have higher content validity and reliability (by reducing measurement error). Multiple questionnaire measures, though, can increase participant and staff burden and may increase study costs.

Some of the more commonly used generic HRQL instruments are the SF-36 [27], the EQ-5D [28], the Rotterdam Symptom Checklist [35], and the Memorial Symptom Assessment Scale [36]. The National Institutes of Health sponsored Patient-Reported Outcomes Measurement Information System (PROMIS) is also a good resource for HRQL and symptom assessment measures [37], and offers options to tailor the measures to meet specific investigator and study needs. Generic pediatric measures include the PedsQL [38], the KidsSCREEN [39], and PROMIS [37].

Frequently used condition-specific instruments include the Functional Assessment of Cancer Therapy (FACT) [40] and the European Organization for Research and Treatment of Cancer Quality of Life (EORTC QLQ) [41], both of which are multidimensional measures assessing the HRQL of individuals with cancer. Other condition specific instruments include the Centers for Epidemiological Studies—Depression (CES-D) [42], the Profile of Mood States (POMS) [43], and the Patient Health Questionnaire (PHQ) [22], all of which assess psychological distress and well-being; and the Barthel Index, which measures physical functioning and independence [44]. There are several good reviews of HRQL measures in the literature [45, 46], as well as on select websites [47].

Within a specific symptom or dimension of HRQL, like physical functioning, one can assess the degree to which an individual is able to perform a particular task, his or her satisfaction with the level of performance, the importance to him or her of performing the task, or the frequency with which the task is performed. Thus, the aspects of HRQL or symptoms measured in clinical trials vary depending on the specific research questions of the trial. When selecting appropriate HRQL instruments, one should consider the specific aspects of the disease/condition or symptom.

Some professional societies advocate the use of certain assessment tools or the measurement of specific sets of symptoms, so study results can be compared across trials using the same measures [48]. Consulting professional societies affiliated with certain conditions or diseases is advisable. For example, the American Society of Clinical Oncology has guidelines for the screening, assessment and care of anxiety and depressive symptoms in adults with cancer [49]. It recommended several screening instruments for ongoing use, with the hope that more uniformity in tracking these symptoms would be established in the cancer area.

Scoring of HRQL Measures

Instruments may be used to assess changes in specific dimensions or symptoms, describe the intervention and control groups at specific times, and examine the correspondence between HRQL measures and clinical or physiological measures. Plans for data analysis are tailored to the specific goals and research questions of the clinical trial.

Most established instruments have standard scoring algorithms. Adhering to these scoring methods is critical in order to interpret scores accurately and compare trial results with those from other studies. In many clinical trials, several measures are used, such that several distinct scores will be calculated (e.g., depression or pain). Some HRQL instruments may also produce an overall HRQL score in addition to separate scores for each HRQL dimension [40].

Determining the Significance of HRQL Measures

An important issue in evaluating HRQL measures is determining how to interpret score changes on a given scale. For example, how many points must one increase or decrease on a scale for that change to be considered clinically meaningful? Does the change in score reflect a small, moderate, or large improvement or deterioration in a participant's health status? Recent years have seen an increase in research examining the question of the clinical significance of HRQL and symptom scores. Demonstrating clinical significance is also important for achieving successful product claims through regulatory agencies [50].

Information on how to interpret changes in HRQL is based on the minimal important difference [51, 52]. When the change in score is connected to clinical measures, the difference is sometimes referred to as the minimal clinically important difference. This difference is defined as the smallest score or change in scores that is perceived by participants as improving or decreasing their HRQL and which would lead a clinician to consider a change in treatment or follow-up [52, 53]. The responsiveness of a HRQL instrument (i.e., the instrument's ability to measure change) and the minimal important difference can vary based on population and contextual characteristics. Thus, there will not be a single value for a HRQL instrument across all uses and populations, but rather a range in estimates that vary across patient populations and observational and clinical trial applications [51]. A variety of methods have been used to determine the minimal important difference. However, there is currently no consensus on which method is best and therefore multiple approaches are used [51, 53, 54]. More in-depth discussion of issues regarding the minimal important difference and HRQL and other PRO measures can be found elsewhere [51].

Utility Measures/Preference Scaling and Comparative Effectiveness Research

The types of HRQL instruments discussed in this chapter have been limited to measures that were derived using psychometric methods. These methods examine the reliability, validity, and responsiveness of instruments. Other approaches to measuring quality of life and health states are used, however, and include utility measures and preference scaling [55, 56]. Utility measures are derived from economic and decision theory, and incorporate the preferences of individuals for particular interventions and health outcomes. Utility scores reflect a person's preferences and values for specific health states and allow morbidity and mortality changes to be combined into a single weighted measure, called quality-adjusted life years (QALYs). These measures provide a single summary score representing the net change in quality of life (the gains from the intervention minus adverse effects and burden). Utility scores are most often used in cost-effectiveness analyses that combine quality of life and duration of life [57–59]. Ratios of cost per QALY can be used to decide among competing interventions.

In utility approaches, one or more scaling methods are used to assign a numerical value from 0.0 (death) to 1.0 (full health) to indicate an individual's quality of life. Procedures commonly used to generate utilities are lottery or standard gamble (most usually the risk of death one would be willing to take to improve a state of health) [56]. Preferences for health states are generated from the general population, clinicians, or patients using multi-attribute scales, visual analogue rating scales, time trade-off (how many months or years of life one would be willing to give up in exchange for a better health state), or other scaling methods [55, 60]. Utility measures

are useful in decision-making regarding competing treatments and/or for the allocation of limited resources. They also can be used as a predictor of future health events. For example, Clarke and colleagues examined the use of index scores based on the EQ-5D, a 5-item generic health status measure, as an independent predictor of vascular events, other major complications and mortality in people with type 2 diabetes, as well as to quantify the relationship between these scores and future survival [61]. The investigators enrolled 7,348 people from Australia and New Zealand, aged 50–75, to the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) study. After adjusting for standard risk factors, a 0.1 higher index score derived from the EQ-5D was associated with an additional 7% lower risk of vascular events, a 13% lower risk of complications, and a 14% lower rate of all-cause mortality. Thus, the EQ-5D was an independent marker for mortality, future vascular events, and other complications in participants with type 2 diabetes.

In general, psychometric and utility-based methods measure different components of health. The two approaches result in different yet related, and complementary assessments of health outcomes, and both are useful in clinical research. Issues regarding the use of utility methods include the methodologies used to derive the valuation of health states; the cognitive complexity of the measurement task; potential population and contextual effects on utility values; and analysis and interpretation of utility data [55, 56]. For a further review of issues related to utility analyses/preference scaling, and the relationship between psychometric and utility-based approaches to the measurement of life quality, additional references may be consulted [55–60, 62].

References

1. Quality of Life Assessment in Cancer Clinical Trials. Report of the Workshop on Quality of Life Research in Cancer Clinical Trials. *USDHHS*, Bethesda, Maryland, 1991.
2. Spilker B (ed.). *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelphia: Lippincott-Raven Publishers, 1996.
3. Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials (Interdisciplinary Statistics)*. Boca Raton, Florida: Chapter & Hall/CRC, 2002.
4. Fayers P, Machin D. *Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes*. Chichester: John Wiley & Sons, Ltd., 2007.
5. Snyder CF, Jensen RE, Segal JB, Wu AW. Patient-Reported Outcomes (PROs): Putting the patient perspective in patient-centered outcomes research. *Med Care* 2013;51:S73-S79.
6. Calvert M, Blazeby J, Altman DG, Revicki DA, et al. for the CONSORT PRO Group. Reporting of Patient-Reported Outcomes in randomized trials. The CONSORT PRO extension. *JAMA* 2013;309:814-822.
7. Calvert M, Brundage M, Jacobsen PB, et al. The CONSORT Patient-Reported Outcome (PRO) extension: Implications for clinical trials and practice. *Health Qual Life Outcomes* 2013;11:184-190.
8. Food and Drug Administration. Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Silver Spring, MD: Office of Communications, Division of Drug Information Center for Drug Evaluation and Research Food and Drug Administration, 2009.

9. Berzon R, Hays RD, Shumaker SA. International use, application and performance of health-related quality of life instruments. *Qual Life Res* 1993;2:367-368.
10. Mossey JM, Shapiro E. Self-rated health: A predictor of mortality among the elderly. *Am J Public Health* 1982;72:800-808.
11. Kaplan GA, Camacho T. Perceived health and mortality: A nine-year follow-up of the human population laboratory cohort. *Am J Epidemiol* 1983;117:292-304.
12. Oei TP, McAlinden NM, Cruwys T. Exploring mechanisms of change: The relationships between cognitions, symptoms, and quality of life over the course of group cognitive-behaviour therapy. *J Affect Disord* 2014;168:C72-C77.
13. Schron EB, Shumaker SA. The integration of health quality of life in clinical research: Experiences from cardiovascular clinical trials. *Prog Cardiovasc Nurs* 1992;7:21-28.
14. Sugarbaker PH, Barofsky I, Rosenberg SA, Gianola FJ. Quality of life assessment of patients in extremity sarcoma clinical trials. *Surgery* 1982;91:17-23.
15. The Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Control Clin Trials* 1998;19:61-109.
16. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *JAMA* 2002;288:321-333.
17. The Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy. *JAMA* 2004;291:1701-1712.
18. Naughton MJ, Jones AS, Shumaker SA. When practices, promises, profits, and policies outpace hard evidence: The post-menopausal hormone debate. *J Soc Issues* 2005;61:159-179.
19. Hays J, Ockene JK, Brunner RL, et al for the Women's Health Initiative Investigators. Effects of estrogen plus progestin on health-related quality of life. *N Engl J Med* 2003;348:1839-1854.
20. Brunner RL, Gass M, Aragaki A, et al for the Women's Health Initiative Investigators. Effects of conjugated equine estrogen on health-related quality of life in postmenopausal women with hysterectomy: results from the Women's Health Initiative Randomized Clinical Trial. *Arch Intern Med* 2005;165:1976-1986.
21. Richardson LP, Ludman E, McCauley E, et al. Collaborative care for adolescents with depression in primary care: a randomized clinical trial. *JAMA* 2014;312:809-816.
22. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606-613.
23. Poznanski E, Mokros H. *Children's Depression Rating Scale-Revised (CDRS-R)*. Los Angeles, CA: WPS, 1996.
24. Bird HR, Andrews H, Schwab-Stone M, et al. Global measures of impairment for epidemiologic and clinical use with children and adolescents. *Int J Methods Psychiatr Res* 1996;6:295-307.
25. Brittenden J, Cotton SE, Elders A, et al. A randomized trial comparing treatments for varicose veins. *N Engl J Med* 2014;371:1218-1227.
26. Garratt AM, Macdonald LM, Ruta DA, et al. Towards measurement of outcome for patients with varicose veins. *Qual Health Care* 1993;2:5-10.
27. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). 1. Conceptual framework and item selection. *Med Care* 1992;30:473-483.
28. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33:337-343.
29. Shumaker SA, Anderson R, Berzon R, Hayes R (eds.) International use, application and performance of health-related quality of life measures. *Qual Life Res* 1993;2:367-368.
30. Cella DF, Wiklund I, Shumaker SA, et al. Integrating health-related quality of life into cross-national clinical trials. *Qual Life Res* 1993;2:433-440.
31. Naughton MJ, Shumaker SA, Anderson R, Czajkowski S. Psychological Aspects of Health-Related Quality of Life Measurement: Tests and Scales. In Spilker B (ed.), *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelphia: Lippincott-Raven Publishers, 1996.
32. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013;22:1889-1905.

33. Wu AW, Bradford AN, Velanovich V, et al. Clinician's checklist for reading and using an article about patient-reported outcomes. *Mayo Clin Proc* 2014;89:653-661.
34. Hays RD, Revicki DA. Reliability and validity (including responsiveness). In: Fayers P, Hays R (eds.). *Assessing quality of life in clinical trials (2nd edition)*. New York: Oxford University Press, 2005.
35. Hardy JR, Edmonds P, Turner R, et al. The use of the Rotterdam Symptom Checklist in palliative care. *J Pain Symptom Manage* 1999;18:79-84.
36. Portenoy RK, Thaler HT, Kornblith AB, et al. The Memorial Symptom Assessment Scale: An instrument for the evaluation of symptom prevalence, characteristics and distress. *Eur J Cancer* 1994;30:1326-1336.
37. National Institutes of Health, Patient-Reported Outcomes Measurement Information Systems website: www.nihPROMIS.org.
38. Varni JW, Seid M, Kurtin PS. PedsQLTM 4.0: Reliability and validity of The Pediatric Quality of Life Inventory™ version 4.0 Generic Core Scales in healthy and patient populations. *Med Care* 2001;39:800-812.
39. Ravens-Sieberer U, Gosch A, Rajmil L, et al. KIDSCREEN-52 quality-of-life measure for children and adolescents. *Expert Rev Pharmacoecon Outcomes Res* 2005;5:353-364.
40. Cella DF, Tulskey DS, Gray G, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;11:570-579.
41. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365-376.
42. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Appl Psych Meas* 1977;1:385-401.
43. McNair DM, Lorr M, Droppleman LF. *Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service, 1981.
44. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: A reliability study. *Int Disabil Stud* 1988;10:61-63.
45. Wilkin D, Hallam L, Doggett M. *Measures of Need and Outcome for Primary Health Care*. New York, NY: Oxford Medical Publications, 1992.
46. McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press, 2006.
47. Website for the International Society for Quality of Life Research (ISOQOL): www.ISOQOL.org
48. Reeve BB, Mitchell SA, Dueck A, et al. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. *J Natl Cancer Inst* 2014;106.
49. Anderson BL, DeRubeis RJ, Berman BS. Screening, assessment, and care of anxiety and depressive symptoms in adults with cancer: An American Society of Clinical Oncology guideline adaptation. *J Clin Oncol* 2014;32:1605-1619.
50. Revicki DA, Osoba D, Fairclough D, et al. Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Qual Life Res* 2000;9:887-900.
51. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-109.
52. Jaeschke R, Singer J, Guyatt G. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1991;12:S266-S269.
53. Guyatt G, Walter S, Norman G. Measuring change over time: Assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171-178.
54. Guyatt G, Osoba D, Wu AW. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002;77:371-383.
55. Weinstein MC, Torrance G, McGuire A. QALYs: The basics. *Value Health* 2009;12:S5-S9.

56. Revicki DA, Kaplan RM. Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. *Qual Life Res* 1993;2:477-487.
57. Neumann PJ, Auerbach HR, Cohen JT, Greenberg D. "Low-value" services in value-based insurance design. *Am J Manag Care* 2010;16:280-286.
58. Greenberg D, Rosen AB, Wacht O, et al. A bibliometric review of cost-effectiveness analysis in the economic and medical literature, 1976-2007. *Medical Decis Making* 2010;30:320-327.
59. Greenberg D, Earle CC, Fang CH, et al. When is cancer care cost-effective? A systematic overview of cost-utility studies in oncology. *J Natl Cancer Inst* 2010;102:82-88.
60. Kaplan RM, Feeny D, Revicki DA. Methods for assessing relative importance in preference based outcome measures. *Qual Life Res* 1993;2:467-475.
61. Clark PM, Hayes AJ, Glasziou PG, et al. Using the EQ-5D index score as a predictor of outcomes in patients with type 2 diabetes. *Med Care* 2009;47:61-68.
62. Moving the QALY forward: Building a pragmatic road. *Value Health* 2009;12:S1-S39