Measurement Issues in Quality-of-Life Assessments

Claudette G. Varricchio, RN, DSN, FAAN

Purpose/Objectives: To describe issues to be considered when measuring the quality of life (QOL) of patients with cancer.

Data Sources: Published articles and books, conference proceedings, and abstracts.

Data Synthesis: A review of the existing literature reveals an emerging field of research and considerable discussion of measurement issues. Consensus is growing about measurement approaches to this subjective concept; however, agreement differs regarding specific aspects of measurement approaches and interpretation of data.

Conclusions: The purpose of the QOL assessment and how the data will be analyzed are the main determinants of the choice of assessment and measurement approaches. Differences regarding how QOL is measured may be encountered based on a clinical or research purpose.

Implications for Nursing: QOL assessments may be used as part of data collection to address a specific research question or may be used to guide clinical practice. Research findings used to guide clinical practice should be evaluated for validity, reliability, and fit of the sample before being incorporated into clinical practice. QOL assessment in clinical practice may be used as an exploratory tool to identify potential problems or may be used to evaluate the effectiveness of a targeted intervention.

Key Points . . .

- ➤ Significant advances have occurred since the mid-1990s in the assessment and measurement of cancer-related quality of life.
- ➤ The science of quality-of-life assessment is evolving; however, no gold standard currently exists for measurement methodology.
- ➤ The purpose of an assessment will determine the appropriate methods to be used.

Quality-of-life (QOL) measurement very often is included in overall assessments as clinicians and researchers, especially in cancer care, strive to better patient outcomes when improvement in disease state may be limited with current therapies or the risk of untoward effects of experimental therapies may be high. To contribute to improved patient outcomes, QOL assessments for research or clinical purposes must be based on sound methods, use reliable and valid approaches, and have findings that are valid and consistent with the measurement approach. This article will discuss the methodologic components that contribute to valid and useful QOL assessment.

QOL assessment has become a central concept in clinical trials and clinical practice. In 2000, a survey of Oncology Nursing Society members identified QOL as the second most important research priority for the organization (Ropka et al., 2002). In a more recent survey to determine the Society's research priorities for 2005–2008, QOL was the most important priority (Berger et al., 2005). In 1995, the Oncology Nursing Society convened a State-of-the-Knowledge Conference to address QOL issues from theoretical, research, clinical, and educational perspectives (King et al., 1997). The group identified definitional and methodologic issues that must be considered and resolved before QOL assessments could be included among the standard assessments that lead to clinical decisions. Other groups have examined the measurement issues related to QOL assessments that are barriers to the adoption of QOL assessment as a standard of care (Mayo Clinic, 2002). Some of the identified barriers are gaps in language and communication between the research literature and clinicians, the absence of unified guidelines for the interpretation of QOL assessments, the availability of numerous instruments without consensus regarding which to use, the addition of a QOL measure to patient assessments without attention to scientific methods, and the perception of QOL assessment as an added burden without added value.

The measurement issues related to QOL assessments are associated with the complexity of the concept. Less agreement exists regarding the exact definition of what constitutes QOL for an individual (Chauhan, Eppard, & Perroti, 2004; Ware, 2003) (see Figure 1). The lack of consensus may be because QOL is an evolving phenomenon. As experts have learned more about QOL, its concepts and descriptors have changed. Additionally, QOL must be considered in the context of the healthcare experience (e.g., disease, treatment). The World Health Organization (1993) defined QOL as a state of complete physical, mental, and social well-being, not merely the absence of disease and infirmity. Five dimensions or domains of the concept generally have been agreed on in the literature: physical functioning or well-being, psychological well-being, social role functioning or well-being, disease- and treatment-related symptoms, and spiritual well-being (Ferrans, 1990a, 1990b; Haberman &

Claudette G. Varricchio, RN, DSN, FAAN, is a retired consultant in Rockville, MD. *(Submitted May 2005. Accepted for publication July 20, 2005.)*

Digital Object Identifier: 10.1188/06.ONF.S1.13-21

- Includes psychological and social function, as well as physical functioning, and incorporates positive aspects of well-being and negative aspects of disease and infirmity (World Health Organization, 1993)
- A person's sense of well-being that stems from satisfaction or dissatisfaction with the areas in life that are important to him or her (Ferrans & Powers, 1985)
- A person's appraisal of and satisfaction with his or her current level of functioning compared to what he or she perceives to be possible or ideal (Cella & Cherin, 1988)
- The term "quality of life" is not simply a catchword for an area of scientific investigation but defines interactions with the physical, emotional, intellectual, and spiritual environments (Chauhan et al., 2004).

Figure 1. Examples of Accepted Definitions of Quality of Life

Bush, 2003; King et al., 1997; King & Hinds, 2003). Most QOL instruments measure all domains or dimensions, but in some instances, the purpose of the assessment may be focused on one domain alone (e.g., the effect of an intervention to control a symptom such as pain or nausea and vomiting). As a result, the researcher or clinician may choose to assess the physical domain more specifically in the context of an overall QOL assessment. In other situations, the effect of the interaction may be assessed regarding the physical and psychological domain on overall QOL; therefore, a measure of pain, or other relevant symptom or outcome, may be added to a measure of overall QOL to give more specific information on which to base the interpretation of effects of an intervention. The assessment of interaction effects will require an approach that allows domain subscores as well as an overall QOL score.

Many decisions must be made when planning to measure QOL or any of its components for research or clinical purposes. This article will provide an overview of current issues and research to facilitate more complete discussions of the topics. Measurement issues to be considered in the assessment of QOL for research or clinical practice are included in Figure 2. Some of the issues that healthcare providers and researchers must address include the (a) purpose of the QOL assessment, (b) main concept of interest, (c) reliability and validity of the measure, (d) generic or disease-specific nature of the measure, and (e) frequency of measurement.

What Is the Purpose of the Assessment?

The purpose for assessing QOL must be determined to drive the measurement and interpretation approaches used in the clinical or research setting. Padilla, Frank-Stromborg, and Koresawa (2004) noted five purposes for assessing QOL: to (a) describe patient responses to disease, (b) describe patient responses to symptom management, (c) describe patient and family responses to cancer treatment, (d) describe patient responses to rehabilitation efforts, and (e) address trajectories in the course of a patient's disease experience. Also, knowing whether the purpose is to develop group data in the context of research or to gather data about an individual to make clinical decisions for that patient is helpful (Cella, Bullinger, Scott, Barofsky, & Clinical Significance Consensus Meeting Group, 2002). These considerations may influence the choice of a single- or multi-item scale, a generic or disease-specific measure, a unidimensional or multidimensional measure of QOL, or a single dimension or influence on QOL (Sloan, Cella, et al., 2002) (see Figure 3).

Measurement choices and interpretation are influenced by the purpose of the measurement exercise, whether it is research on group effects or individual assessment and intervention (Sloan, Aaronson, et al., 2002; Sloan, Cella, et al., 2002). If the purpose of a clinical assessment is to measure QOL before an intervention and, after the evaluation, to determine the effect of an intervention rather than to assess QOL to determine the need for an intervention, different assessment methods may be used (Haberman & Bush, 2003; King & Hinds, 2003; Symonds, Berzon, Marquis, Rummans, & Clinical Significance Consensus Meeting Group, 2002). The purpose of assessing QOL should be clear when researchers design a study. Clinicians and researchers may want to ask what type of disease, condition, or treatment will be assessed; what type of QOL tool will be used and whether it is reliable and valid; and what the objective of the research is (i.e., to evaluate treatment effects or identify QOL problems) (Rubenstein, 1996; Symonds et al.). An extensive discussion of these issues can be found in King and Hinds.

The research or clinical focus of a QOL evaluation will play a significant role in the choice of the measurement approach, which could be framed as group versus individual assessments. An assessment as part of a research project will require greater rigor and perhaps more complexity than a clinical assessment. A research assessment may intend to establish group data or norms, whereas a clinical assessment most likely will aim to determine what is happening to an individual and how the status of that individual compares to a group with known data on the effect of the same variable on QOL (Cella et al., 2002). A research assessment may appear to be more extensive and rigorous than a clinical assessment, which may have a more pragmatic focus and use a single item as a probe or screening approach to determine whether additional, more specific assessment is needed. A single dimension of QOL may be a reasonable follow-up to a response

- What is the purpose of the assessment?
- What is the conceptual fit between the research or clinical question and the proposed measure?
- Is the assessment for research or clinical evaluation?
- Are existing valid and reliable measures of this concept available?
- What kind of scale or measure should be used?
  - Is a generic measure, a disease-specific measure, or a combination of both appropriate to assess the variable of interest?
- What are the characteristics of the measure?
  - Is a valid comparison group available to permit a meaningful interpretation of the scores?
  - If the original measure was validated on an English-speaking sample, is the translation a validated, conceptual, and linguistically equivalent form?
  - How often should patients be assessed?
- What can affect the analysis and interpretation of data?
  - Is respondent burden an issue?
  - How will missing data be considered in the analysis?
  - Is response shift a factor?
  - Do clinical and statistical significance exist?
  - Does the measure have information about what constitutes a clinically meaningful change in the score?

Figure 2. Questions to Guide Decisions About Measurement Approaches

Disease-specific measure: provides additional information on a single dimension or subpopulation of cases such as a specific kind of cancer or a specific set of expected outcomes

Generic measure: assesses quality of life over a range of situations and provides a metric for comparison of quality of life among heterogeneous populations

Multidimensional scale: measures more than one dimension of quality of life and may yield a composite or specific domain score

Multi-item scale: a questionnaire composed of items that are conceptually related and may be aggregated into a collective score

Single-item scale: a single question that can stand alone or be part of a series of loosely affiliated questions or part of a psychometrically sound measurement index

Figure 3. Categories of Quality-of-Life Measures

Note. Based on information from Haberman & Bush, 2003; Sloan, Aaronson, et al., 2002.

to a screening or assessment question. To be useful in clinical practice, a measure or tool must be valid, reliable, and easy to use and interpret and should not be a burden to clinicians or patients. An evaluation as part of the research must meet all of the criteria for validity, reliability, and generalizability.

What Is the Conceptual Fit Between the Research or Clinical Question and the Proposed Measure?

The first consideration in choosing a measure of QOL is to clearly identify what will be measured. What is the main concept of interest? Is the construct to be measured as a global overview of the individual's QOL, limited to health-related QOL, or limited to one or more dimension of QOL (Ferrans, 1990b; Guyatt et al., 2002; Vallerand & Payne, 2003)?

Conceptual fit is the degree to which the measure assesses the concepts or variables that interest the researcher or clinician. The conceptual focus of a measure is reported by its developer in publications that describe its psychometric properties. The research or clinical question must determine how the assessment will be made (King et al., 1997; Sloan, Aaronson, et al., 2002). Then, the question will be used to determine the method of assessment, means of assessment, and possible interpretation. A conceptual fit must exist between the means of assessment and the purpose of the assessment. In other words, was the measure designed to assess the same concept addressed in the purpose? If researchers are interested in how a situation (disease or symptom) affects functional status, do the measures ask about the disease, symptom status, progression, and aspects of function (either physical or role function) and QOL? A measure of functional status cannot be used to assess overall QOL or directly assess a specific symptom included in one of the QOL domains. If the purpose is to determine whether an intervention is affecting pain as one dimension of QOL, the measure must include the level of pain as a minimum along with a measure of QOL. Other aspects of QOL believed to be influenced by pain (e.g., functional status) may be included in the assessment. A global assessment of QOL will not give direct information about the pain level or the effect of an intervention to manage pain. The assessment of a specific symptom or dimension will not provide an overall evaluation of the effect of pain on QOL unless a generic measure of QOL is included. Therefore, a clear identification of the purpose of the assessment and how the information will be used is needed to determine whether a conceptual fit exists with the factors included in the assessment tool.

In addition to clarifying the concepts and conceptual fit, the definition being used to guide the QOL assessment should be explained, whether the assessment is being used for research or clinical practice. Defining QOL to individuals being assessed is essential so that they understand how to respond to the assessment; for example, the person being assessed must know the time frame being used (i.e., today, last week, since the last assessment).

Is the Assessment for Research or Clinical Evaluation?

Although researchers are interested in statistical significance, clinicians usually are concerned with the benefit for or effect on a single individual (i.e., is the difference large enough to have implications for patient care?) (Sloan, Cella, et al., 2002). Researchers generally report group data that often are compared with normative data from previous studies in groups with similar characteristics. The clinical use of QOL data frequently is related to clinical decision making for an individual (Cella et al., 2002).

In clinical practice, data from an individual might be used to describe the person's health state, screen for disease, assess the individual's needs, monitor disease progression, or monitor treatment response (Cella et al., 2002; McHorney & Tarlov, 1995). An extensive discussion of clinical significance of group versus individual data can be found in an article by Cella et al. (2002).

Choice of a Measurement Approach

Are Existing and Reliable Measures of the Concept Available?

Existing validated measures are used for many reasons, including establishing psychometric parameters, possibly establishing normative group data for comparisons, comparing findings across studies using the same measures, and eliminating the tedious work of developing a new measure. If an existing measure and the research question at hand conceptually fit, the use of an existing measure is recommended. A listing of commonly used scales with the descriptions and psychometric properties of each can be found in Padilla et al. (2004) and Omery and Dean (2004). Figure 4 lists frequently used QOL assessment tools; in addition, Web-based resources for existing and new measures, with comments about psychometric and descriptive information, appear in Figure 5.

Reliability and validity are the most important aspects to evaluate when choosing a tool to measure QOL. Validity is the degree to which the instrument measures what it purports to measure. Reliability is the degree to which the tool consistently measures what it purports to measure. If researchers are evaluating a new area of study in which no validated measure of the concept exists, they should seek guidance from research texts and consultation with experts in measure development. The process of instrument development and validation is complex and should not be attempted by researchers who do

Generic Questionnaires

The Beck Depression Inventory (BDI)
The Crumbaugh Purpose-in-Life Test (PIL)
Demands of Illness Inventory (DOII)
Functional Assessment of Cancer Therapy–General Scale (FACT-G)
Global Adjustment to Illness Scale
Lewis Psychological Coherence Scale
McCorkle & Young Symptom Distress Scale
McGill Pain Questionnaire
McMaster Health Index Questionnaire
Medical Outcome Study–Short Form General Health Survey (MOS)
The Norbeck Social Support Scale
The Nottingham Health Profile
Profile of Mood States (POMS)
Psychosocial Adjustment to Medical Illness (PAIS)
Quality of Life Index by Padilla et al. (QLI)
Quality of Life Index by Spitzer et al. (QL-Index)
The Rosenberg Self-Esteem Scale
Sickness Impact Profile (SIP)
Spielberger State-Trait Anxiety Inventory (STAI)
Ware Health Perceptions Questionnaire

Cancer-Specific Questionnaires

Breast Cancer Chemotherapy Questionnaire
Bush Bone Marrow Transplant Symptom Inventory
Cancer Rehabilitation Evaluation System (CARES)
City of Hope, Quality of Life: Bone Marrow Transplant Demands of Bone Marrow Transplant Inventory (DBMT)
European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ-C30). Modules for lung cancer and bone marrow transplant symptoms
Functional Assessment of Cancer Therapy Scale (FACT-G). Modules for head and neck, breast, prostate, lung, and colorectal cancer
Functional Living Index: Cancer (FLIC)
Linear Analog Self-Assessment (LASA): Breast Cancer
Quality Adjusted Time Without Symptoms or Toxicity (Q-TWIST)
Quality of Life Index: Cancer Version, by Ferrans and Powers
Rotterdam Symptom Checklist (QOL after breast cancer)
Southwest Oncology Group Quality of Life Questionnaire
Time Without Symptoms or Toxicity (TWIST)

Figure 4. Examples of Frequently Used Quality-of-Life Questionnaires

Note. From "Quality of Life Methodological and Measurement Issues" (p. 181), by M.R. Haberman & N. Bush in C.R. King & P.S. Hinds (Eds.), *Quality of Life From Nursing and Patient Perspectives: Theory, Research, Practice* (2nd ed.), 2003, Sudbury, MA: Jones and Bartlett. Copyright 2003 by Jones and Bartlett. Reprinted with permission.

not have specific training and expertise in this area (Hays, Anderson, & Revicki, 1998).

What Kind of Scale or Measure Should Be Used?

An item is defined as a single question that can stand alone or be part of a series of loosely affiliated questions or part of a psychometrically sound measurement index (Sloan, Aaronson, et al., 2002). A single question or item may be used in some situations in which more global or less precise information is sufficient. For example, on the Spitzer Uniscale, an individual is asked, "Please rate your overall QOL," and responds by marking an X on a line from 0–100 (Sloan et al., 1998). The single item from the European Organisation for Research and Treatment of Cancer QOL Questionnaire-30 asks, "How would you rate your overall QOL during the past week?" (Aaronson et al., 1993). A single-item measure may be appropriate for screening that will lead to a more detailed assessment if a problem is indicated by a patient's score. A multi-item scale is a questionnaire composed of items that are related conceptually and may be aggregated into a collective score. Multi-item questionnaires can measure various aspects of a single domain of QOL, such as physical function, or a single symptom. Researchers and clinicians should differentiate between a multi-item scale and a multidimensional scale (Haberman & Bush, 2003) and between an assessment of QOL and a symptom assessment that is not part of a QOL assessment.

Some QOL experts recommend using multiple tools to measure QOL versus a single tool, which may be helpful when evaluating multiple dimensions or can provide flexibility in the conceptualization of QOL. This method also can allow for comparability and responsiveness by using a battery of tools, composite tools constructed from portions of existing instruments, or a core tool with specific modules (e.g., European Organisation for Research and Treatment of Cancer QOL Questionnaire-30, Functional Assessment of Cancer Therapy scale) (Omery & Dean, 2004).

A questionnaire may be divided into multiple domains or dimensions assessing more than one aspect of QOL with corresponding subscores. A multi-item scale measures one variable or dimension using many questions. A multidimensional scale measures more than one dimension of QOL and may yield a composite of or specific domain scores. Most of the more well-known QOL tools are multidimensional scales that attempt to measure the domains of a construct and generate a subscore for each domain. The subscores may be reported separately or summed to give a single score for the questionnaire.

The purpose of an assessment should guide the choice of a single- versus a multi-item measure. A single-item index may be more appropriate if a tool is needed for description or screening. If an assessment is needed to determine the need for a specific intervention or to assess the effectiveness of an intervention, a multi-item index will give more detailed information about the status and contributing factors so that

Child Health Questionnaire: www.healthact.com/chq (measures for children ages 5 years and older)

Cochrane Collaboration: www.cochrane.org/reviews/mainindex.html (systematic reviews of the effect of healthcare interventions)

Functional Assessment of Chronic Illness Therapy: www.facit.org (measurement system information)

Health Measurement Research Group: www.healthmeasurement.org (information about approaches to health management)

International Society for Quality of Life Research: www.isoqol.org (resources and links to sites for measures and annotated bibliographies of favorite and classic articles)

Mapi Research Institute: www.mapi-research.fr (cultural adaptations and linguistic validation of questionnaires)

PedsQL: www.pedsql.org (PedsQL™ 4.0 measurement model)

Quality of Life Instruments Database: www.qlmed.org (thousands of existing quality-of-life measures with detailed information on more than 300 tools)

Figure 5. Web Resources for Identification and Evaluation of Quality-of-Life Measures

a more meaningful interpretation may be made. The purpose of the assessment also will determine whether a single score will provide the information required or whether the added information of results on specific subscales will provide better guidance in interpreting the findings (Haberman & Bush, 2003; Sloan, Aaronson, et al., 2002).

QOL scores may be used to describe the responses of a group in a research endeavor. Scores usually are reported as a mean for the group studied, for a specific variable assessed, or for the group on a single- or multi-item measure. Scores often are compared to a group mean achieved for a similar normative group in the process of validation of the measure. This approach is useful in evaluating the overall effect of a variable on a group (Cella et al., 2002; Redelmeier & Tversky, 1990).

Generic Versus Disease-Specific Scales

Generic QOL measures assess QOL over a range of situations and provide a metric for comparison of QOL over heterogeneous populations. Disease-specific measures are tailored to assess the specific symptoms and effects known to be associated with a condition or treatment (Dijkers, 2003; Padilla et al., 2004). The Medical Outcomes Study–Short Form (Ware, Kosinski, & Keller, 1994) and the QOL Index by Spitzer et al. (1981) are examples of generic measures that can be used across populations to assess QOL. Normative data from the general population or specific disease groups are available for comparison.

When disease-specific information is required, an additional questionnaire or single-dimension scales can be added to the assessment. For example, a group of disease- and symptom-specific scales was developed as an adjunct to the Functional Assessment of Cancer Therapy Scale–General (Cella et al., 1993). The disease-specific modules include head and neck, breast, prostate, lung, and colorectal cancer.

The purpose of an assessment and how the information will be used to guide research or clinical decisions will determine which approach should be taken. When deciding how QOL will be assessed and how comprehensive an assessment should be, the issue of respondent burden must be evaluated. If respondent burden is too great, it is a threat to the reliability of the data generated and may bias the findings.

What Are the Characteristics of the Measure?

Is a Valid Comparison Group Available?

Normative fit is the degree to which the group being evaluated matches the description of the group on which the measure was validated. For example, when assessing a group of older adults with prostate cancer, the measure used should have been validated on a similar group, not a group of young patients or a group with heart disease. To some degree, valid interpretation is related to the ability to interpret the scores or findings of a comparison group. The group may have characteristics similar to the general population or to patients with a common disease state. Researchers must know the characteristics of the group in which the measurement tool was developed and on which its validity and reliability were established. The characteristics of the normative group also should match the characteristics of the sample or individual being assessed (Cella et al., 2002; Kendall, Marrs-Garcia, Nath, & Sheldrick, 1999).

Researchers would expect to get very different information from a measure used on a healthy population and an unhealthy population. If the purpose of an assessment is to determine how the scores of individuals with cancer compare to a healthy population, then the chosen measure should have information on scores for the healthy group (e.g., group norms). The scores of the individuals with cancer being assessed would be expected to differ from those of the healthy normative group. However, if the purpose is to compare the effect of an intervention to improve symptoms in a group of patients, a measure should have comparison data from a group of patients who have the same symptom and probably the same disease but have not received the intervention being tested (Guyatt et al., 2002).

The research question or clinical use of the data may require that the age of the normative group be considered, which is especially true when assessing children or adults who are in different developmental stages that could influence how they respond to questions or an intervention (Hymovich, 2004; Rasin, 2004). Researchers should consider how closely the characteristics of the groups on which the measure was validated match the group to be assessed. A measure may be valid and reliable in a population of adults with one disease and set of characteristics but may be invalid and unreliable among those in another age group or with another disease. Similarly, a measure that is valid in assessing responses to treatment may not be valid when used with cancer survivors or in prevention settings. The more appropriate norm group for survivors or in prevention settings may be a healthy population rather than patients currently receiving treatment. The fit of the measure with the characteristics of the group being examined may be a direct threat to the validity of the study or the clinical application of the findings of a valid study. Most QOL assessment measures are validated in an adult population similar to those who meet the inclusion criteria for cancer clinical trials. Specific measures for use with children, older adults, patients in hospice care, or other special populations are being developed but are not commonly available (Hymovich; McMillan & Small, 2002; McMillan & Weitzner, 1998; Rasin).

Is the Translation a Valid, Conceptual, and Linguistically Equivalent Form?

The issues around translations and adaptations of existing measures to be used in non–English-speaking groups or languages other than the original have been addressed in numerous publications (Berry, 1980; Montero, 1977; Varricchio, 2004). Guidelines offer recommendations for the translation of existing measures and discussions of conceptual versus linguistic equivalence in the literature (Brislin, 1970, 1986; Marin & Marin, 1991). A rigorous process should be followed to achieve valid translation of items (see Figure 6). Moreover, the original tool should use simple sentences, an active voice, no metaphors or slang, specific rather than general terms, and no vague terms, such as "probably" and "maybe."

Conceptual equivalence occurs when the items measure the same concepts in both languages. In the process of establishing equivalence, researchers may determine that no exact word conveys the same concept in both languages. As a re-

- Be familiar with the content (e.g., disease vocabulary, concepts) in the source language and in the target language.
- Use words in the source language that have similar frequency of use in the target language.
- Translators and back-translators should work independently of each other.
- Test the translation on bilinguals.
- Refine translations on items where there is ambiguity or discrepancy in responses.
- Discard items where agreement on the wording or meaning cannot be achieved. Modification of the wording of items in the source language may be necessary at this point.
- Test with focus groups or a small pilot group of the target population to ensure that persons representative of the target group understand the items.
- Administer the items to bilingual subjects: Some see the source language version, some see the target language version, and some see both. Responses should be similar across groups.

Figure 6. Recommendations for Linguistic and Conceptual Translation of Existing Measures: Back-Translation Method

Note. From "Measurement Issues Concerning Linguistic Translations" (p. 56), by C.G. Varricchio in *Instruments for Clinical Health-Care Research* (3rd ed.), by M. Frank-Stromborg & S.J. Olsen (Eds.), 2004, Sudbury, MA: Jones and Bartlett. Copyright 2004 by Jones and Bartlett. Reprinted with permission.

sult, the author must explore the concept in the new language to discover a thought that conveys the same meaning. This exercise may lead to clarification of the original measure's language or to the discovery that the new language does not have an equivalent concept. The process of establishing conceptual equivalence in two languages is as rigorous as the development of a completely new measure (Varricchio, 2004; Warnecke et al., 1996).

Linguistic equivalence is a direct translation using words that are as close as possible to the original items. The preferred method is back translation, which begins with one individual translating a tool from the source language (e.g., English) to the target language (e.g., Spanish). Then, a second person translates the target version back to the source version. The second version is compared for equivalence to the original. This process does not explore whether the translated item conveys the same concept as the original item, which may lead to different responses in the two language groups when the items no longer ask about the same concepts. Ultimately, a flawed interpretation of the findings may result. A related issue is the use of measures developed and validated on adults to assess children, older adults, or other special populations, including those at the end of life, or the use of measures developed to assess treatment responses of survivors or those in prevention studies (Varricchio, 2004).

How Often Should Patients Be Assessed?

When and how often to assess patients are design issues that should be determined by the research question or clinical usefulness. The rate of variability of the concept being measured also influences the timing and frequency. If the variable of interest is subject to frequent change in response to an intervention or other influences, more frequent assessments may be warranted (e.g., longitudinal evaluations over a specific time period). If a stable situation is expected, a cross section, or a one-time assessment at a specific point, may be preferred. When determining the frequency of assessments, the expected rate of change (i.e., time points at which the change is expected) and the burden on the subject of multiple assessments must be considered (Sprangers et al., 2002). Many QOL experts believe an individual's QOL is a state that can vary over time and, if measured only once, the true QOL picture may not emerge. Yet, a cross-sectional approach does allow for comparisons among individuals at one time point (Padilla et al., 2004). A list of variables to be considered in the process of determining when to measure aspects of QOL is presented in Table 1. When choosing a measure, its known responsiveness or sensitivity in detecting minimally significant changes must be evaluated (Patrick & Chiang, 2000). Responsiveness is the ability of a measure to detect minimally important changes in specific situations. The importance of the magnitude of change depends on the baseline status of the individual or group. Many measures of QOL have either floor or ceiling effects in specific populations, meaning that if a person scores very low or very high, the measure will not pick up small changes that show the individual is doing better or worse (Sprangers et al., 2002).

Arguments can be made for assessing only the points that will be used in an analysis. Economic considerations in terms of time and effort by researchers, clinicians, and participants may be important. Some investigators maintain that regular assessments coinciding with clinic visits or other encounters with the researchers or clinicians reduce missing data. As with all design decisions, the research question or clinical use of the data will determine the best approach (Sprangers et al., 2002).

Issues Affecting Analysis and Interpretation of Data

Is Respondent Burden an Issue?

Patient or respondent burden might affect the response rate and completeness of data in a research evaluation. If patients tire, they may not answer all of the questions, contributing to the problem of missing data. Another possibility is that respondents will tire and answer falsely or randomly. The order of the measures presented to patients also must be considered as a possible source of respondent burden.

How Will Missing Data Be Considered?

Missing data are an important threat to the valid interpretation of the findings of QOL assessments. The literature addressing this point in the context of research is significant (Fairclough, 2004; Fong, Lam, Kwan, Sham, & Karlberg, 2004; Moinpour et al., 2000; Sloan & Varricchio, 2001). Missing data points cannot be ignored when analyzing and interpreting data, especially in QOL assessments. When and why the data are missing could change the analysis and interpretation of the findings. If data are missing because a subject was too ill to provide data or died before the study ended, the interpretation of the study will be very different than if the missing data points are ignored in the analysis or if they are assigned dummy values that could influence the outcome of the analysis. Therefore, every effort should be made to collect all data on all subjects included in an analysis. In addition, reasons for missing data and subject exclusion in the analysis should be documented. Selective or random dropping of data from a subject may lead to false interpretation of the findings (Fairclough, 1997).

Table 1. Variables That Affect Selection of Time Points for Quality-of-Life (QOL) Assessment

Variable	Example Rationale
Baseline assessments mandatory	Cannot measure change without an assessment before the initiation of treatment
Natural course of the disease	Known points of remission and worsening; indolent disease results in slower changes than rapidly progressive disease.
Disease stage: early	Longer follow-up period to address survivorship issues, monitor late effects (both positive and negative), and see if patients are able to return to "normal" activities
Disease stage: late	Shorter follow-up period because of patient status and potential for missing data
Effects associated with the treatment course or administration	Documentation of acute, short-term adverse effects or cumulative adverse effects (e.g., at the end of radiotherapy)
Timing of clinical event monitoring	To coincide with, for example, tumor size measurements
Timing of important clinical events	Assessing QOL at progression may warrant patient-specific measurement times for patients with the event
Completion of treatment and/or a short time after completion of treatment	Resolution of mucositis may require 2–4 weeks after completion of radiotherapy. Treatment arms might be compared to the end of radiotherapy and 2–4 weeks later to see how much better or sooner palliation occurs.
Respondent burden	Too many assessments confer patient burden and affect adherence with the QOL assessment schedule.
Data collection and management resources	Frequent assessments require more personnel and data management effort.
Data collection before administration of treatment and/or discussion with clinical staff	Want to compare patient experience with different regimens after recovery from previous cycle. Also want to avoid biasing patient report based on feedback from medical staff.
Timing of QOL assessment should be similar across treatment arms.	Becomes a problem when regimens have different administration schedules (e.g., 3-week, 4-week cycles); the choice of assessments can be based on time (e.g., every second week following baseline) or on event (e.g., every second treatment cycle)

Note. From "Assessing Meaningful Changes in Quality of Life Over Time: A Users' Guide for Clinicians," by M.A. Sprangers et al., 2002, *Mayo Clinic Proceedings, 77,* p. 564. Copyright 2002 by the Mayo Foundation. Reprinted with permission.

Can Response Shift Account for the Findings?

Response shift is the change in scores over time as respondents adjust their expectations to allow for variations in appraisal because of differences in health status or expectations (Sprangers et al., 2002; Westerman & Hak, 2004). In some situations, a response shift can lead to data that are very difficult to interpret or to erroneous interpretations (Sprangers et al., 1999; Sprangers & Schwartz, 1999). Response shift may account for findings that are counter to expectations based on the clinical state of the patient. The phenomenon of response shift often is observed in longitudinal studies or repeated measures over time when the observed clinical state of individuals seems to disagree with their scores on QOL assessments. Patients may adjust their expectations as they progress through the disease trajectory from newly diagnosed and rather healthy to treatment and on to the end of life (Sprangers et al., 2002; Westerman & Hak). QOL may not be as bad or as good as patients expected it to be. Another explanation that has been offered is that, to maintain hope, a person will readjust his or her evaluation of his or her state. This response shift may be reflected in rather flat scores or improvement in scores in the presence of clinical deterioration.

Do Clinical and Statistical Significance Exist?

Statistical significance is the degree of confidence that the findings in a study did not occur randomly. Reports of data analysis using specific statistical measures include a preset significance level (i.e., p values) that assists readers in determining the likelihood that the tested intervention was related to the outcome of interest. Design issues, such as sample size and frequency of assessment, can influence the statistical significance of findings. Many reports have revealed that statistical significance may have no relationship to clinical significance (Jaeschke, Singer, & Guyatt, 1989; Kazdin, 1999; Osoba, Rodrigues, Myles, Zee, & Pater, 1998; Sloan, Aaronson, et al., 2002). Clinical significance goes beyond statistical significance to determine whether the difference is large enough to have implications for clinical care (Sloan, Cella, et al., 2002). In research settings, the definition of a clinically meaningful change in score is determined during the design phase of the study (i.e., a priori), before any analysis of the data is undertaken. The degree of change may be determined from clinical observation or from the analysis of the results of many studies using the same measure. This approach determines the amount of change in the score that constitutes a meaningful clinical difference. An extensive discussion of this topic can be found in Guyatt et al. (2002).

Clinical significance may differ based on the perspective of the person evaluating the data. Researchers, clinicians, patients, and the general public may have very different ideas of what is clinically significant. Researchers may report statistical significance for the group studied, whereas clinicians may evaluate whether the patient's condition has improved or deteriorated in a way that would require a change in the clinical management of the situation. Patients may be interested in the amount of change and whether that change is seen as important. The public or population perspective may include the degree to which the data might be used to determine the health state of a community or as an influence on public policy (Frost et al., 2002). When using information to extrapolate the effect on an individual, the variability of individual responses must be considered. If the data are used as a basis for individual clinical decisions, a clinically meaningful change in score must be defined for the measure. Group norms describe how the group under study scored in the situation. To extrapolate this information to an individual in clinical practice, the characteristics of the individual versus

those in the research sample must be evaluated. Furthermore, the general expectation of clinical benefit in the situation must be understood (Guyatt et al., 2002).

Does the Measure Have Information About a Clinically Meaningful Difference?

A concept related to clinical and statistical significance is that of a clinically meaningful significance or a minimally clinically important difference (MCID), which is the smallest change in a QOL score considered to be worthwhile or clinically important (Hays & Woolley, 2000; Symonds et al., 2002). The clinically meaningful difference change in score for most QOL measures is not known. For many years, researchers believed that small differences in QOL might be statistically significant but clinically unimportant (Wyrwich et al., 2005). The concept of MCID was introduced and the question raised as to what the MCID is for any specific QOL tool. However, problems were encountered in using the MCID. First, the MCID is derived from the average change in QOL for a group, not an individual. In addition, the amount of change might depend on the direction (i.e., positive or negative, getting better or worse) or the meaning of change might depend on baseline (the starting point). The goal of researchers is to identify clinically meaningful differences in the interpretation of the QOL scores or the assessment. Although this goal may be worthwhile for researchers and clinicians to understand the measures being used, it still is fallible (Hays & Woolley). Interest has increased in establishing information on clinically meaningful differences for the most common measures (Osoba et al., 1998; Samsa et al., 1999). The lack of information makes interpretation or comparison of findings from clinical trials difficult (Guyatt et al., 2002).

Conclusion

Measurement of QOL, in all of its forms, can be a positive tool in improving clinical outcomes; however, all of the measurement issues addressed in this article must be considered by researchers and clinicians. Clinicians must be familiar with the psychometrics of a measure and the characteristics of the population studied before deciding to incorporate the measure or study findings into practice. Clinicians and researchers must have a good understanding of what was measured and how to interpret scores before making judgments or recommendations for clinical practice based on the scores. QOL data are useful as a basis for clinical decisions only if they can be relied on to be valid. Inappropriate choices of measures yield data that are not useful and may impede good clinical practice.

Author Contact: Claudette G. Varricchio, RN, DSN, FAAN, can be reached at varricchio@comcast.net, with copy to editor at ONFEditor@ons.org.

References

Aaronson, N.K., Admedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N.J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute, 85,* 365–376.

Berger, A.M., Berry, D.L., Christopher, K.A., Greene, A.L., Maliski, S., Swenson, K.K., et al. (2005). Oncology Nursing Society year 2004 research priorities survey. *Oncology Nursing Forum, 32,* 281–290.

Berry, J. (1980). Acculturation as varieties of adaptation. In A.M. Padilla (Ed.), *Acculturation: Theory, models, and some new findings* (pp. 9–25). Boulder, CO: Westview Press.

Brislin, R.W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1,* 185–216.

Brislin, R.W. (1986). The wording and translation of research instruments. In W.J. Lonner & J.W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Beverly Hills, CA: Sage.

Cella, D., Bullinger, M., Scott, C., Barofsky, I., & Clinical Significance Consensus Meeting Group. (2002). Group vs individual approaches to understanding clinical significance of differences or changes in quality of life. *Mayo Clinic Proceedings, 77,* 384–392.

Cella, D.F., & Cherin, E.A. (1988). Quality of life during and after cancer treatment. *Comprehensive Therapy, 14*(5), 69–75.

Cella, D.F., Tulsky, D.S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., et al. (1993). The Functional Assessment of Cancer Therapy scale: Development and validation of the general measure. *Journal of Clinical Oncology, 11,* 570–579.

Chauhan, C., Eppard, W., & Perroti, J. (2004). The patient advocate perspective on assessing the clinical significance for quality-of-life measures. *Journal of Cancer Integrative Medicine, 2,* 155–157.

Dijkers, M.P. (2003). Individualization in quality of life measurement: Instruments and approaches. *Archives of Physical Medicine and Rehabilitation, 84*(4, Suppl. 2), S3–S14.

Fairclough, D.L. (1997). Summary measures and statistics for comparison of quality of life in a clinical trial of cancer therapy. *Statistics in Medicine, 16,* 1197–1209.

Fairclough, D.L. (2004). Patient reported outcomes as endpoints in medical research. *Statistical Methods in Medical Research, 13,* 115–138.

Ferrans, C.E. (1990a). Development of a quality of life index for patients with cancer. *Oncology Nursing Forum, 17*(3, Suppl.), 15–19.

Ferrans, C.E. (1990b). Quality of life: Conceptual issues. *Seminars in Oncology Nursing, 6,* 248–254.

Ferrans, C.E., & Powers, M.J. (1985). Quality of life index: Development and psychometric properties. *Advances in Nursing Science, 8*(1), 15–24.

Fong, D., Lam, K.F., Kwan, C.W., Sham, J.S., & Karlberg, J.P. (2004). False positive error in the analysis of paired quality of life data [Abstract 1330]. *Quality of Life Research, 13,* 1521.

Frost, M.H., Bonomi, A.E., Ferrans, C., Wong, G.Y., Hays, R.D., & Clinical Significance Consensus Meeting Group. (2002). Patient, clinician, and population perspectives on determining the clinical significance of quality-of-life scores. *Mayo Clinic Proceedings, 77,* 488–494.

Guyatt, G.H., Osoba, D., Wu, A.W., Wyrwich, K.W., Norman, G.R., & Clinical Significance Consensus Meeting Group. (2002). Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings, 77,* 371–383.

Haberman, M.R., & Bush, N. (2003). Quality of life methodological and measurement issues. In C.R. King & P.S. Hinds (Eds.), *Quality of life from nursing and patient perspectives: Theory, research, practice* (2nd ed., pp. 171–198). Sudbury, MA: Jones and Bartlett.

Hays, R.D., Anderson, R.T., & Revicki, D. (1998). Assessing reliability and validity of measurement in clinical trials. In M.J. Staquet, R.D. Hays, & P.M. Fayers (Eds.), *Quality of life assessment in clinical trials: Methods and practice* (pp. 169–182). Oxford, England: Oxford University Press.

Hays, R.D., & Woolley, J.M. (2000). The concept of clinically meaningful difference in health-related quality-of-life research: How meaningful is it? *Pharmacoeconomics, 18,* 419–423.

Hymovich, D.P. (2004). Measurement issues with children and adolescents. In M. Frank-Stromborg & S.J. Olsen (Eds.), *Instruments for clinical health-care research* (3rd ed., pp. 33–46). Sudbury, MA: Jones and Bartlett.

Jaeschke, R., Singer, J., & Guyatt, G.H. (1989). Measurement of health status.

Ascertaining the minimal clinically important difference. *Controlled Clinical Trials, 10,* 407–415.

Kazdin, A.E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting Clinical Psychology, 67,* 332–339.

Kendall, P.C., Marrs-Garcia, A., Nath, S.R., & Sheldrick, R.C. (1999). Normative comparisons for the evaluation of clinical significance. *Journal of Consulting and Clinical Psychology, 67,* 285–299.

King, C.R., Haberman, M., Berry, D.L., Bush, N., Butler, L., Dow, K.H., et al. (1997). Quality of life and the cancer experience: The state-of-the-knowledge. *Oncology Nursing Forum, 24,* 27–41.

King, C.R., & Hinds, P.S. (Eds.). (2003). *Quality of life from nursing and patient perspectives: Theory, research, practice* (2nd ed.). Sudbury, MA: Jones and Bartlett.

Marin, G., & Marin, B.V. (1991). *Research with Hispanic populations.* Newbury Park, CA: Sage.

Mayo Clinic. (2002). Symposium on the clinical significance of quality of life measures in cancer patients. *Mayo Clinic Proceedings, 77*(Suppl.), 367–563.

McHorney, C.A., & Tarlov, A.R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research, 4,* 293–307.

McMillan, S.C., & Small, B.J. (2002). Symptom distress and quality of life in patients with cancer newly admitted to hospice home care. *Oncology Nursing Forum, 29,* 1421–1428.

McMillan, S.C., & Weitzner, M. (1998). Quality of life in cancer patients: Use of a revised Hospice Index. *Cancer Practice, 6,* 282–288.

Moinpour, C.M., Sawyers Triplett, J., McKnight, B., Lovato, L.C., Upchurch, C., Leichman, C.G., et al. (2000). Challenges posted by non-random missing quality of life data in an advanced-stage colorectal cancer clinical trial. *Psycho-Oncology, 9,* 340–354.

Montero, D. (1977). Research among racial and cultural minorities: An overview. *Journal of Social Issues, 33*(4), 1–10.

Omery, A.K., & Dean, H. (2004). Multiple instruments for measuring quality of life. In M. Frank-Stromborg & S.J. Olsen (Eds.), *Instruments for clinical health-care research* (3rd ed., pp. 150–161). Sudbury, MA: Jones and Bartlett.

Osoba, D., Rodrigues, G., Myles, J., Zee, B., & Pater, J. (1998). Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology, 16,* 139–144.

Padilla, G.V., Frank-Stromborg, M., & Koresawa, S. (2004). Single instruments for measuring quality of life. In M. Frank-Stromborg & S.J. Olsen (Eds.), *Instruments for clinical health-care research* (3rd ed., pp. 128–163). Sudbury, MA: Jones and Bartlett.

Patrick, D.L., & Chiang, Y.P. (2000). Measurement of health outcomes in treatment effectiveness evaluations: Conceptual and methodological challenges. *Medical Care, 38*(9, Suppl.), II14–II25.

Rasin, J.H. (2004). Measurement issues with the elderly. In M. Frank-Stromborg & S.J. Olsen (Eds.), *Instruments for clinical health-care research* (3rd ed., pp. 47–55). Sudbury, MA: Jones and Bartlett.

Redelmeier, D.A., & Tversky, A. (1990). Discrepancy between medical decisions for individual patients and for groups. *New England Journal of Medicine, 322,* 1162–1164.

Ropka, M.E., Guterbock, T., Krebs, L., Murphy-Ende, K., Stetz, K., Summers, B., et al. (2002). Year 2000 Oncology Nursing Society research priorities survey. *Oncology Nursing Forum, 29,* 481–491.

Rubenstein, L.V. (1996). Using quality of life tests for the patient diagnosis or screening or to evaluate treatment. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed., pp. 363–374). Philadelphia: Lippincott-Raven.

Samsa, G., Edelman, D., Rothman, M.L., Williams, G.R., Lipscomb, J., & Matchar, D. (1999). Determining clinically important differences in health status measures: A general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics, 15,* 141–155.

Sloan, J.A., Aaronson, N., Cappelleri, J.C., Fairclough, D.L., Varricchio, C.G., & Clinical Significance Consensus Meeting Group. (2002). Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic Proceedings, 77,* 479–487.

Sloan, J.A., Cella, D., Frost, M., Guyatt, G.H., Sprangers, M., Symonds, T., et al. (2002). Assessing clinical significance in measuring oncology patient quality of life: Introduction to the symposium, content overview, and definition of terms. *Mayo Clinic Proceedings, 77,* 367–370.

Sloan, J.A., Loprinzi, C.L., Kuross, S.A., Miser, A.W., O'Fallon, J.R., Mahoney, M.R., et al. (1998). Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. *Journal of Clinical Oncology, 16,* 3662–3673.

Sloan, J.A., & Varricchio, C. (2001). Quality of life endpoints in prostate chemoprevention trials. *Urology, 57*(4, Suppl. 1), 235–240.

Spitzer, W.O., Dobson, A.J., Hall, J., Chesterman, E., Levi, J., Shepherd, R., et al. (1981). Measuring the quality of life of cancer patients: A concise QL index for use by physicians. *Journal of Chronic Diseases, 34,* 585–597.

Sprangers, M.A., Moinpour, C.M., Moynihan, T.J., Patrick, D.L., Revicki, D.A., & Clinical Significance Consensus Meeting Group. (2002). Assessing meaningful changes in quality of life over time: A users' guide for clinicians. *Mayo Clinic Proceedings, 77,* 561–571.

Sprangers, M.A., & Schwartz, C.E. (1999). The challenge of response shift for quality-of-life-based clinical oncology research. *Annals of Oncology, 10,* 747–749.

Sprangers, M.A., van Dam, F.S., Broersen, J., Lodder, L., Wever, L., Visser, M.R., et al. (1999). Revealing response shift in longitudinal research on fatigue—The use of the thentest approach. *Acta Oncologica, 38,* 709–718.

Symonds, T., Berzon, R., Marquis, P., Rummans, T.A., & Clinical Significance Consensus Meeting Group. (2002). The clinical significance of quality of life results: Practical considerations for specific audiences. *Mayo Clinic Proceedings, 77,* 572–583.

Vallerand, A.H., & Payne, J.K. (2003). Theories and conceptual models to guide quality of life related research. In C.R. King & P.S. Hinds (Eds.), *Quality of life from nursing and patient perspectives: Theory, research, practice* (2nd ed., pp. 45–64). Sudbury, MA: Jones and Bartlett.

Varricchio, C.G. (2004). Measurement issues concerning linguistic translations. In M. Frank-Stromborg & S.J. Olsen (Eds.), *Instruments for clinical health-care research* (3rd ed., pp. 56–64). Sudbury, MA: Jones and Bartlett.

Ware, J.E., Jr. (2003). Conceptualization and measurement of health-related quality of life: Comments on an evolving field. *Archives of Physical Medicine and Rehabilitation, 84*(4, Suppl. 2), S43–S51.

Ware, J.E., Jr., Kosinski, M., & Keller, S.D. (1994). *SF 36 physical and mental health summary scales: A user's manual.* Boston: New England Medical Center.

Warnecke, R.B., Ferrans, C.E., Johnson, T.P., Chapa-Resendez, G., O'Rourke, D.P., Chavez, N., et al. (1996). Measuring quality of life in culturally diverse populations. *Journal of the National Cancer Institute Monographs, 20,* 29–38.

Westerman, M.J., & Hak, T. (2004). Response shift in quality of life in patients with small cell lung cancer [Abstract 1196]. *Quality of Life Research, 13,* 1521.

World Health Organization. (1993). *WHO QOL study protocol: The development of the World Health Organization quality of life assessment instrument* (Publication No. MNH/PSF/93.9). Geneva, Switzerland: Author.

Wyrwich, K.W., Bullinger, M., Aaronson, N., Hays, R.D., Patrick, D.L., Symonds, T., et al. (2005). Estimating clinically significant differences in quality of life outcomes. *Quality of Life Research, 14,* 185–295.