

Patient-Reported Outcomes

1. Learning Objectives

After reviewing this chapter readers should be able to:

- Define patient-reported outcomes and their role in research and practice;
- Identify different types of patient-reported outcome measures and their strengths and weaknesses in application;
- List and give examples of the concepts and domains incorporated into health status, quality of life, and health-related quality of life measures;
- Identify and define the measurement properties of new or existing patient-reported outcome measures, with special attention to validity of the assessment; and
- Write a critical appraisal of a patient-reported outcome assessment in a clinical trial.



2. Introduction

While you and i have lips and voices which
are for kissing and to sing with
who cares if some one-eyed son of a bitch
invents an instrument to measure Spring with?

--e. e. cummings, "voices to voices, lip to lip... (XXXIII)"

The extract from the poem above illustrates his distinctive use of words, his rejection of common rules for writing, such as capitalization and punctuation, his irreverence, and his originality. He challenges us to think about the wisdom and goal of measuring quality of life or what many in the field are beginning to call *Patient-Reported Outcomes (PROs)*, because quality of life has many broad meanings and using PRO terminology requires specification of the concepts being measured. PROs in the poem are analogous to the measurement of 'spring'.

From Poetry to PROs

- Why should people care about measuring the patient's view and voice?
- Are there any limits to our attempts, i.e., can anything about experience of health and illness be assigned a number that clearly means one person's health state is better than another's or better today than yesterday?
- What is it about PROs that can be measured and best can we go about this task?
- How do we label what it is we are trying to measure most appropriately?



Exercise 1: Level of Importance

Below are a selection of common things individuals find important in life. Where they fall in importance is impacted, however, by the health status of the individual. Consider the importance of these factors in your own life. Then, using your mouse, drag and drop these factors into the spaces provided to create your own list of importance for when you are well and when you are ill.



Exercise



Paying Bills



Working



Seeing Friends



Eating Well

When Well	Level of Importance	When Ill
	High	
	Low	

3. Reasons to Measure PROs

There are several important features of self reported measures in medicine and public health:

- They are used increasingly to help determine whether treatments are doing more good than harm;
- These outcomes are assessed and often compared to clinical measurements that remain the primary end-points for most clinical trials and for many clinicians, because they are familiar through long or repeated use; and
- Epidemiological investigations and population surveys incorporate self-reported outcomes to compare populations and to describe the status of different populations. Sometimes these are called *quality of life indicators*, although more frequently the term *health status indicators* best describes the content of these measures.

Difference Between Clinical Measures and Self-Report Measures

Clinical measures include physiologic measures that require professional knowledge to interpret and clinician judgments that come from interviews and observations of patients.

Self-report measures of health and quality of life often have more meaning to the persons who are affected by disease, are undergoing treatments, or are trying to restore or maintain health.

Because perceptions of health and illness influence what people do about their health (e.g. visit doctors, go to a hospital, or ignore signs and symptoms), policy makers are also increasingly interested in self-reported outcomes.

Self-reported measures differ from clinical measures in three important ways.

1. Self-report measures often correlate poorly with physiologic measures.
Example: Exercise capacity in the laboratory is only weakly related to exercise capacity in daily life (Guyatt et al., 1985).

2. Another common disconnect is the observation that two patients with the same clinical status or physiologic state may have dramatically different responses to the condition.
Example: It is not uncommon for two patients experiencing back pain with the same range of motion and even similar ratings of back pain to have different abilities to work and different responses to pain. One patient with back pain may stay home from work, while another with the same pain rating goes snow skiing. In patients with major depression, depending on individual and environmental factors, some patients may continue to work and take care of their families, while others quit their jobs and stay at home.
3. Perhaps most importantly, assumptions by clinicians and investigators that physiologic outcomes predict those outcomes that are important measures of benefit often turns out to be incorrect (Kunz et al., 2008).

3. Reasons to Measure PROs



Exercise 2: Patient-Important Outcomes

For each surrogate in the following table, identify the patient-important outcomes (PIOs) with which there is a putative causal link. When you make a correct match a summary of results of randomized trials will appear. These results disappointed investigators by showing that their understanding of biology may be terribly flawed: compelling biological rationale failed to predict impact on patient-important outcomes, sometimes disastrously so.

Drag the following outcomes into the PIO column below which you think was the trial result.

▶ Death Dyspnea and Fatigue Renal Failure QOL impairment, coronary events QOL impairment, vertebral and longbone fractures ◀

Intervention	Surrogate Endpoint	Patient-Important Outcomes
Milrinone, amrinone, ibopamine, epoprotonol, vesnarinone for patients with heart failure	Measures of cardiac function	
Hormone replacement therapy for post-menopausal women	High-Density Lipoprotein and ultrasound-measured vessel wall thickening	
Fluoride for women with osteoporosis	Bone density	
Encainide and flecainide for cardiac arrhythmias	Arrhythmia suppression	
Rosiglitazone for diabetes	Serum glucose or Hemoglobin A1C	
Torcetrapib for prevention of coronary events	HDL	
Nitric oxide for critically ill patients with acute respiratory distress syndrome (ARDS)	Oxygenation	
Theophylline for asthma	Spirometric measures of lung function	

4. Individualizing Quality of Life Measures

Each person is first and foremost an individual, a somebody, a someone, a soul, a human being. Thus, it is not a large leap to define quality of life as “what the person or patient says it is” (Joyce et al., 1988).

Emphasis on the individual differs from most commonly-known quality of life measures, which reject the idiographic view that one can find comparable cues to what quality of life is between individuals, let alone between groups or cultures. The domains of quality of life in the idiographic view must, instead, be chosen by the individual himself or herself and then evaluated in relation to the individual's environment, such that the person and the environment intersect (Joyce et al., 1999).



Example 1: Observational Studies

A 25 year old athlete is engaged in numerous sports, and does very well. For reasons he finds hard to understand, he finds his life unsatisfying and is generally not very happy. A diving accident results in quadriplegia, leaving him completely dependent. After a month of depression, he finds the meaning in life that had previously eluded him. When asked about his mood, he describes himself as satisfied with his limited life, and overall very happy.

Q: Was his quality of life better before, or after, his accident?

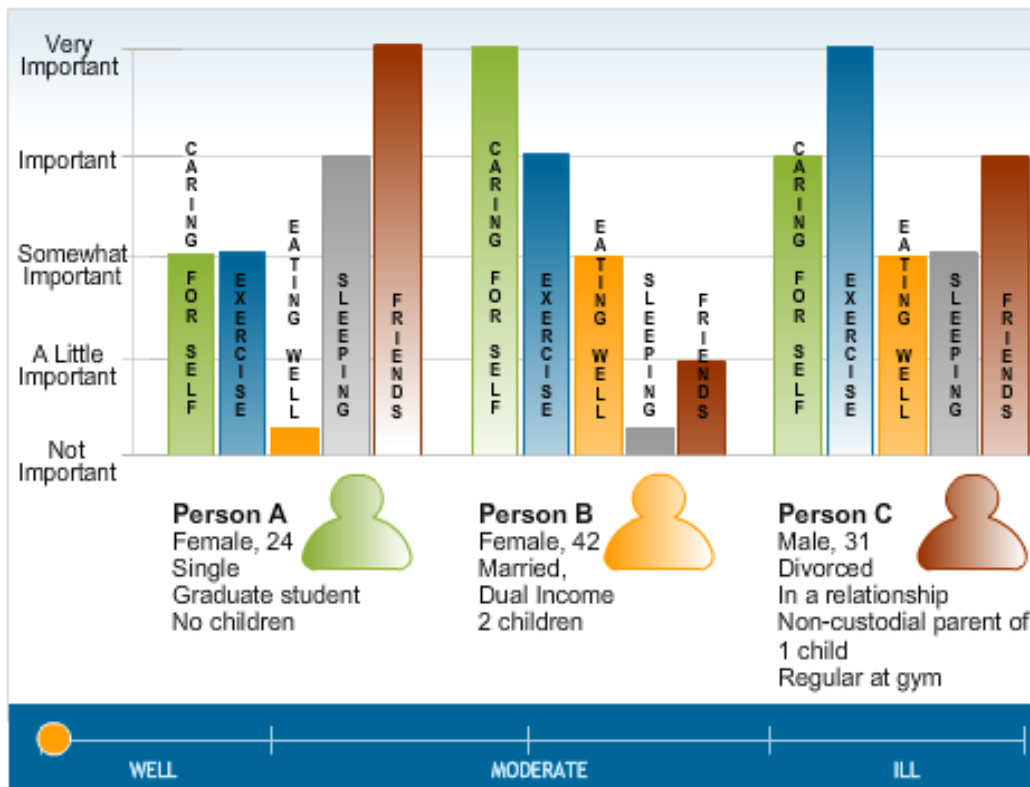
A: If one adopts this perspective strictly, no two measures of quality of life might look the same, as each measure would tap individual experience that may be constrained by the particular environment of the individual or by the moment in time. A scientist who completely rejects the nomothetic ideal of finding a general “law” or “pattern” might well find sympathy with the poem by e.e. cummings. In the end, all individuals have separate identities, physically, emotionally, and spiritually.



Exercise 3: Quality of Life when Ill versus Well

In the illustration below is an example of three individuals and indicators for their personal importance levels for five areas of their lives (Very Important, Important, Somewhat Important, A Little Important or Not Important). Using your mouse, **slide the orange dot** in the bottom bar to different places of health to view how the importance of these areas change depending on whether the person is well or ill.

Ask three individuals you know what they would define as quality of life and how it might change if they were ill versus well.



4. Individualizing Quality of Life Measures

Types of Assessment

Many outcome researchers seek some compromise in the debate between the completely idiographic approach, which emphasizes the individual and the particular, and the nomothetic approach emphasizing the group and the general. There are four well recognized approaches to assessment, described here.

Individualized Assessment:

Some researchers have developed individualized or patient-generated instruments that allow the respondent to choose the aspects of life to be included in the assessment (Joyce et al, 1999). Most generally the methods of assessment are standardized in that response scales are uniform across application and descriptive statistics are used. Although debate has ensued on the readiness and usefulness of these individualized measures in clinical trials (Patel et al, 2003), these measures address the essential notion that quality of life is an individual matter. Perhaps of most use in clinical practice where a doctor and patient negotiate treatments on an individual basis, variants of these individualized approaches will continue to be developed and applied.

Modular Assessment:

Other researchers adopt a modular approach to quality of life assessment that permits cultural variation. This approach is illustrated by the instruments developed by groups originally organized by the World Health Organization—the WHOQOL or World Health Organization Quality of Life Measures (WHOQOL Group, 1993; 1994; 1995; 1996; 1998). In WHOQOL instrument development, the search was made for the nomothetic in themes about quality of life occurring in different languages and cultures. What “bubbles” up to the top was considered to be comparable across cultures, although the importance of the domains in the WHOQOL instruments may vary across cultures (Skevington et al, 2004). In the WHOQOL instrument development, however, each country or culture develops modules that are specific to their interests and concerns. Thus, the instrument goes part way toward a group-specific measure.

Needs Based Assessment:

Another important direction is the needs-based approach to quality-of-life assessment (McKenna and Doward, 2004; Doward and McKenna, 2004; McKenna et al, 2004). This model builds on major advances in needs theory that arose in the 1940s and 1950s from investigations into human motivation (Maslow, 1943). Individuals are viewed as driven or motivated by their needs. The relation between needs and QoL has been explored repeatedly by many investigators in many different contexts (Hornquist, 1982; Patrick et al, 1988). The needs-based model recognizes the importance of deriving the content of an instrument directly from persons with the condition. Content of these measures is developed in relation to reports of patients' experience in satisfying needs. Life is seen as gaining its quality from the ability and capacity of individuals to satisfy their needs, either inborn or learned, and a high quality of life is achieved when an individual satisfies needs.

Modern Test Theory:

Applications of modern test theory--computer adaptive testing, item banking, and item response theory--represent yet another direction toward individualizing health status assessment. The use of item response theory (IRT) to measure self-reported outcomes is growing rapidly. IRT has been used for many years in educational testing to develop achievement tests and entrance exams that relate item difficulty in a test to a person's ability to answer questions correctly. Analogous to probabilistic Guttman scaling, item response theory relates characteristics of items (item parameters) and characteristics of individuals (latent traits) to the probability of a positive response. A variety of IRT models have been developed for dichotomous (i.e., yes/no or true/false) and polytomous data (i.e., excellent, very good, fair, poor). In each case, the probability of endorsing a particular response category can be represented graphically by an item (option) response function (IRF/ORF).

A useful application of IRT in the health field is computer-adaptive testing (CAT), a measurement approach in which the selection of items is tailored for each respondent (Ware et al., 2003; Cook et al., 2005). The development of a CAT requires several steps that are not required in the development of a traditional measure, including identification of *starting* and *stopping* rules. CAT's most attractive advantage is its efficiency. Greater measurement precision can be achieved with fewer items. For example, a patient who is unable to walk would *skip out* of items pertaining to walking and answer only those items related to his or her baseline status.

Item banks contain health status and quality of life items that are *cross-calibrated* (McHorney and Cohen, 2000; Lai et al., 2003). When used in dynamic testing, two individuals could be compared who answer different items because the items have all been put into a measurement system that provides scoring for all items on a similar metric. The National Institutes of Health in the United States is sponsoring a large effort as part of the NIH ROADMAP for Medical Research to develop a Patient-Reported Outcomes Measurement Information System (PROMIS) network. This trans-NIH initiative aims to use IRT, item banks, and computer adaptive testing to measure patient-reported symptoms such as pain and fatigue and aspects of health-related quality of life across a wide variety of chronic diseases and conditions. (See www.nihpromis.org)

4. Individualizing Quality of Life Measures



Exercise 4: Item Response Theory Approach

Which set of questions would be more amenable to an item response theory approach?
Click on the table you think is the correct set of questions.

Table A Question on Vision

Can you see clearly enough to drive at night?
Can you see clearly enough to drive during the day?
Can you read fine print?
Can you read the newspaper?
Can you see clearly watching television?
Does vision make it difficult for you to cook?
Can you recognize people when they are close to you?

Table B Question on Feelings

Do you often feel sad?
Do you often feel anxious?
Are you often worried?
Do you sometimes feel happy?
Do you have difficulty concentrating?
Do you have difficulty sleeping?

Important Considerations in Individualized Quality of Life Assessment

All attempts to “individualize” quality-of-life assessments are important advances to assessment. But however sophisticated may be the methods used in the assessment, some important issues remain:

- What are the boundaries between *health* and *quality of life*?
- What is the universe of content that constitutes quality of life?
- Is it anything the individual says?
- Is it all the needs an individual might invoke when ill or well?
- Are there any boundaries to health?

5. Setting Boundaries: The Terminology of Patient-Reported Outcomes

Health status, functional status, well-being, quality of life, and health-related quality of life are concepts that are often used loosely and interchangeably. There is no consensus and widely adopted definition of *quality of life* because it is used in different contexts by different people. There is considerable agreement, nonetheless, that the quality of life construct is more comprehensive than health status. Health is only one domain.

Quality of Life

Quality of life includes aspects of the environment that may or may not be affected by health or perceived health. Some widely valued aspects of human existence are not generally defined as health status, such as a safe environment, adequate housing, guaranteed income, and freedom.



Exercise



Source: Adapted from Economist Intelligence Unit's "Worldwide quality-of-life index, 2005."



Example 2: Observational Studies

Quality of Life is more than Health

The Intelligence Unit of the widely distributed publication, *The Economist*, recently developed a new “quality of life” index based on a methodology that links the results of subjective life-satisfaction surveys to the objective determinants of quality of life. Using life-satisfaction surveys as a starting point, the Unit used nine indicators that had a significant influence on life-satisfaction and turned these into an equation that explained more than 80% of the variation in country’s life-satisfaction scores. The main factor was income, but the others things were also important: health, freedom, unemployment, family life, climate, political stability and security, gender equality, and family and community life. Note that health is listed among many other aspects of life. Ireland came out top with the fourth-highest GDP per head in the work in 2005, low unemployment, and political liberties. The U.S. was ranked 13th.

These rankings do not match those for infant mortality or life expectancy, but represent the notion that quality of life is broader than health status.

Health-related quality of life (HRQL)

These global human concerns may adversely affect or be affected by health status. Usually, however, health status measures do not include items relating to income, freedom, and respect. One approach to this boundary problem is to use the term *health-related quality of life (HRQL)*. HRQL is often used to indicate that the measure is concentrated on the health concept and the field of health outcomes but includes perceptions or domains not limited to function. No widely-used, specific definition has emerged of HRQL and thus it is difficult to define or bound even this more limited term.

Most often HRQL is defined as physical, emotional, and social functioning. This equation with functional status can be erroneous and of particular concern to persons with disabilities. Persons with functional limitations may enjoy high quality of life through environmental supports or simply through their own life perspective and evaluation of their needs and desires. Although function may be important to many evaluations of their health, health-related quality of life or quality of life should not be used as synonyms.

Definitions

- It is widely agreed that the **health status** concept and its domains and constructs range from negatively valued aspects of life, including death, to the more positively valued aspects such as role function and happiness. Health status is a useful concept in the context of assessing health services and treatment effectiveness.
- **Functional status** measures usually refer to limitations in the performance of social roles or activity limitations. The status concept is highly dependent on the perspective of the assessor and the assessed.
- **Well-being** measures refer to subjective perceptions, including reports of unpleasant or pleasant sensations and global evaluations of health or subjective status. Symptoms may be included in well-being measures or considered separately. Well-being and quality of life may be distinguished by the level of evaluation, i.e., quality of life contains more global evaluations of life position and perspectives, and well-being contains more domain-specific perspectives such as psychological or physical (Kahneman et al., 1999).

5. Setting Boundaries: The Terminology of Patient-Reported Outcomes

The boundaries of concepts and their definition depend upon the measurement objectives, the funding sponsors' motives, the users' particular concerns, and most important of all, the evidence or data on the concept and constructs (Patrick and Bergner, 1990). Investigators may be interested in defining the health of populations to discover or document unmet needs, to determine the effect of medical interventions, or to guide allocation of resources. Traditional measures of morbidity and mortality are limited in defining health status and leave the texture of peoples' lives unexplored.



Example 3: WHOQOL Definition of Quality of Life

The World Health Organization Quality of Life (WHOQOL) group has defined quality of life as:

“individuals' perceptions of their position in life in the context of the culture and value systems in which they live, and in relation to their goals, expectations, standards, and concerns” (The WHOQOL Group, 1994; The WHOQOL Group, 1995; Szabo, 1996).

This definition reflects the growing recognition that quality of life can be inherently subjective, although normative definitions have been proposed that include more objective standards as well as perceptions of objective conditions (Campbell et al., 1976; Calman, 1987).

Quality of life can be used as:

- A descriptor, i.e., the presence or absence of a characteristic of life;
- An evaluative statement, i.e., some value is attached to the characteristics of an individual, population, or kind of human life or;
- A normative or prescriptive statement, i.e., certain norms indicate which characteristics must be present to have a life of quality.

The WHOQOL group places quality of life squarely in the two traditions of an internal psychological and physiological mechanism producing a sense of satisfaction or gratification with life (Hornquist, 1982), and those external conditions that trigger the internal mechanism

(Rogerson, 1995). Thus, quality of life is a **broad ranging concept that incorporates in a complex way individuals' physical health, psychological state, level of independence, social relationships, personal beliefs, and their relationships to salient features of the environment** (The WHOQOL Group, 1994). The measure is definitely health-related, however, and puts health concerns, not surprisingly for a health organization, in central position of the definition.

6. Conclusion: Label What We Measure

Self-Reported Outcomes, referred to as *Patient Reported Outcomes (PROs)* in the context of healthcare, include any report coming directly from the person or persons affected by their life, health condition(s) and treatment (Patrick et al., 2007). PROs:

- Address the source of the report rather than the content;
- Include health status and quality of life; and
- Report on satisfaction with treatment and care, adherence to prescribed regimens when directly related to end-result outcomes, and any other treatment or outcome evaluation obtained directly from patients through interviews, self-completed questionnaires, diaries, or other data collection.

Developers and users should specify and label the content and type of measure for every application of a PRO and provide evidence of its appropriateness to the intended use, for validity of the measure as used in a particular case, and how to interpret results.

Table 1 presents examples of validation of interpretation of PROs. A major challenge faces developers and users of these measures in establishing a testable theory of the expected and observed relationships among the different concepts and domains of quality of life.



Table 1: Validity and Interpretability

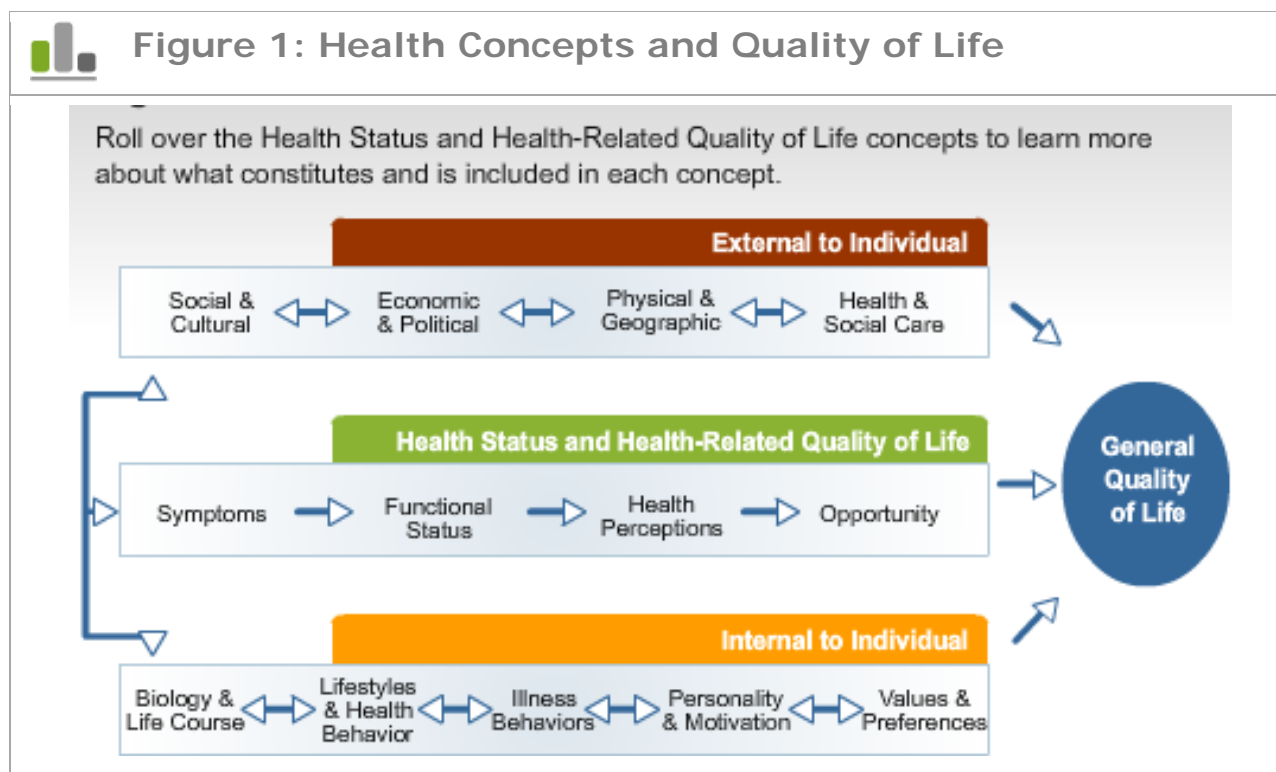
Validity and Interpretability of Patient-Reported Outcomes

Measure	Validation	Interpretation
Measure of dyspnea in patients with lung disease	Correlation with walk test distance, global ratings of dyspnea	An improvement of 4 points represents a small but important difference
Measure of fatigue in patients with heart failure	Correlations with other questionnaires that tap into fatigue	An improvement of 0.5 represents a small but important difference
Measure of vision-related quality of life	Correlation with traditional visual acuity measures	Responses in accord with logical expectations for ordered effects (respondents who can drive at night almost always have no other limitations due to vision; respondents unable to recognize others when they are close are severely limited by vision in all aspects of their lives)

It is as important to establish a theory of how to link clinical variables with health-related quality of life as it is to link larger determinants of PROs such as political unrest, economic depression, inequalities, and sociocultural trends and processes (Wilson and Cleary, 1995; Patrick and Erickson, 1993; Patrick and Chiang, 2000).

6. Conclusion: Label What We Measure

Figure 1 shows one attempt to depict the relationship between *health concepts* and *general quality of life*, and how determinants from the internal (individual) as well as the external (social and cultural) environment influence the general quality of life.



In this figure, health status and health-related quality of life concepts include symptoms, functional status, perceptions, and opportunity (Patrick and Erickson, 1993).

In Figure 1, these are depicted as following a linear progression from the most proximal to the individual, symptoms, to the most distal, opportunity. In reality, no such linear progression exists and symptoms may be translated directly to perceptions or opportunity without affecting function. Most important to the figure is the depiction of the determinants intrinsic to the individual and outside the individual indicating that what people report cannot be separated from the personal or sociocultural environment.

Researchers tend to approach the relationship inductively by collecting data and examining the correlations, but a *priori* hypotheses are important for developing systematic knowledge of how

disease and treatment impacts different indicators of health outcome. The most appropriate approach to causal modeling, the use of health outcomes in meta-analyses, development and application of community-level indicators of health, and interpretation of observed health and quality-of-life measurements remain challenges for both the developers and users of these measures.

PRO terminology permits primary identification that this information comes directly from the patient and avoids confusion in using one or more concepts as an over-arching term with little specification. Some concepts can be measured from both a patient and an observer perspective, e.g., physical function. Others can only be patient-reported, e.g., pain. If a conclusion about pain is based on a non-PRO measure, it would be important to know. The term PRO makes it more likely that use of a PRO is considered in a clinical trial. Perhaps it is reasonable to confirm or add value to an objective or observed finding with a PRO. PROs in a daily diary format may capture daily variations in symptoms or function that can elucidate the mechanism of treatment effects.

Using PRO terminology requires that the concepts be identified. When using this organizing concept, investigators need to define and label the content of a specific PRO. Operational definitions of each construct or concept in use are necessary and recommended, including specification of the theoretical basis for the concept and how the concept was translated into a measurement, i.e. the mathematical definition of assigning a number to a response.

6. Conclusion: Label What We Measure

The concepts and domains included in the measurement of PROs help in making operational definitions. Table 2 contains core concepts and domains contained in many PRO measures.



Table 2: PRO Measures

Concepts and Domains Used in Defining Self-Reported Health Status, Quality of Life, and Health-Related Quality of Life

Concepts	Domains and Attributes
Symptoms +	
	<i>Frequency, severity, bothersomeness</i>
	Reports of physical and psychological symptoms or sensations not directly observable, e.g., energy and fatigue, nausea, irritability
Functional Status +	
	<i>Frequency, difficulty, severity, ability, with or without help</i>
Physical	Functional limitations and activity restrictions, e.g., self care; walking, mobility, and sometimes sleep, and sexual function when construed narrowly
Psychological	Positive or negative affect and cognitive, e.g. anger, alertness, self-esteem, sense of well-being, distress
Social	Engagement, limitations in work, school, play, household management, participation in the "community"
Health Perceptions +	
	<i>Frequency, severity/intensity, satisfaction</i>
Global	General ratings of health and quality of life, e.g., satisfaction or overall well-being
Worries and Concerns	About health, finances, the future

Spiritual +	
	Meaning and purpose of life, connection to a deity, a belief system, or the universe
Disadvantage/Opportunity +	
	<i>Frequency, impact</i>
	Perceptions of stigma or reports of discrimination because of health condition, reports of advantage
Resiliency +	
	<i>Frequency, satisfaction, ability</i>
	Reports of ability to cope or withstand stress and illness
Environmental +	
	<i>Satisfaction, importance</i>
	Evaluations of personal safety, adequacy of housing, respect, freedom, etc.
Satisfaction with Treatment +	
	<i>Expectations, importance, satisfaction</i>
	Reports of treatment and treatment experience
Adherence to Prescribed or Recommended Treatment* +	
	Behaviors directly linked to outcomes Reports of taking treatment, doses, attendance, or Routine behaviors like tooth-brushing intermediate to end results
	*Included not as an end result of treatment but sometimes closely linked causally to treatment and thus a close proxy for a health status outcome. Reports of other behaviors like smoking, alcohol consumption, etc. are generally considerate intermediate and not end-result outcomes though this has been disputed by many.

Table 2 and Figure 1 also demonstrate the multidimensionality of the concepts of health status and health-related quality of life, and as a result require multiple indicators to measure.

6. Conclusion: Label What We Measure



Table 3: Health Status and QOL

Condition	Clinical Signs & Symptoms	Aspects of Life Affected
Acne	An inflammatory skin condition characterized by superficial skin eruptions around hair follicles. Major Symptoms: Skin rash or lesion on the face, neck, chest, shoulders and back / comedones (whiteheads or blackheads) / pustules / cysts / papules / nodules / inflammation around the skin eruptions	<ul style="list-style-type: none"> • Going to social events like dances • Feeling depressed and lonely • Itching, redness, raw skin (symptoms) • Difficulty finding romantic partners
Osteoarthritis	Degeneration of the cartilage that lines the joints. Major Symptoms: pain/tenderness, swelling, creaking, stiffening of affected joints, weakness & shrinkage of surrounding muscles due to lack of use (because of pain), enlarged & distorted joints	<ul style="list-style-type: none"> • Difficulty bending, kneeling, stooping • Difficulty walking up hill • Depressed • Difficulty turning over in bed

Regardless of how items are arranged or how domains are grouped, it is generally agreed that the content validity of PROs can be judged only by the persons or populations being assessed. Thus, the extent to which the domain of interest is comprehensively sampled by the items or questions in the measure can only be judged by representatives of the target population. If the target population is unable to speak for themselves, proxy judgments are sometimes considered acceptable, particularly if supported by rigorously controlled observational studies with inter-rater reliabilities. Before assuming that people cannot speak for themselves, however, they

should be asked and every effort should be made to communicate with them directly. Proxy responses are not PROs.

In addition to content validity, the **other psychometric properties of quality-of-life measures include:**

1. Specification of the measurement model including the instrument's scale and sub-scale structure and the conceptual and empirical basis for combining multiple items into a single score;
2. Reliability, including the degree to which the instrument is free from random error either by testing the homogeneity of content on multi-item tests with internal consistency evaluation or testing the degree to which the instrument yields stable scores over time;
3. Construct, criterion, and predictive validity wherein the logical relationships among different measures are examined;
4. Responsiveness or ability of the measure to assess change over time when real change has occurred (longitudinal construct validation); and
5. Interpretation of the effect size, or the degree to which one can assign qualitative meaning to an instrument's quantitative scores (Scientific Advisory Committee, Medical Outcomes Trust, 2002; Patrick et al., 2007).

Note that rigorous adherence to standards of measurement of PROs is as important as finding the most appropriate label for what is being measured.

7. The Special Case of Validity and PROs

In research studies using PROs special attention needs to be paid to the assessment of validity.

When no gold or criterion standard exists, PRO investigators have borrowed validation strategies from clinical and experimental psychologists who have dealt with the problem of deciding whether questionnaires examining intelligence, attitudes, and emotional function are really measuring what they are supposed to measure.

The most rigorous approach to establishing validity is called *construct validity*. A construct is a theoretically derived notion of the domain(s) we want to measure. An understanding of the construct will lead to expectations about how an instrument should behave if it is valid. The first step in construct validation is to establish a *model* or theoretical framework that represents an understanding of what investigators are trying to measure. That theoretical framework provides a basis for

Definition:

Validity examines whether the instrument is measuring what it is intended to measure.

Types of validity

- *Face validity* examines whether an instrument appears to be measuring what it is intended to measure; and
- *Content validity* examines the extent to which the domain of interest is comprehensively sampled by the items, or questions, in the instrument. Quantitative testing of face and content validity are rarely attempted.
- *Construct validity* involves the logical relations that should exist between two concepts and then comparisons between measures of these concepts to examine if the hypothesized, logical relations are confirmed by the data.

understanding the behavior of the system being studied and allows hypotheses or predictions about how the concepts and instruments being tested should relate to other concepts and their measures. Investigators then administer instruments containing similar and dissimilar concepts to a population of interest and examine the data. Validity is strengthened or weakened when the hypotheses are confirmed or refuted. For example, using a PRO to discriminate between known groups may be validated by comparing two groups of patients: those who received a toxic chemotherapeutic regimen and those who received a less toxic regimen. Any PRO instrument should distinguish between these two groups; if it does not discriminate, something has gone wrong. Alternatively, correlations between symptoms and functional status can be examined; those patients with a greater number and severity of symptoms should have lower functional status scores on a PRO instrument. Another example is the validation of an instrument discriminating between people according to some aspect of emotional function; results should correlate with existing measures of emotional function.



Example 4: Construct Validation

A Detailed Example of Construct Validation

The Inflammatory Bowel Disease Questionnaire (IBDQ) was designed to measure disease-specific HRQL and it includes 30 items directed at 4 domains: bowel symptoms, systemic symptoms, emotional function, and social function. Investigators administered the IBDQ (along with global ratings of change in function, global ratings of change by the physician and a relative, a Disease Activity Index, and the emotional function domain of a generic HRQL measure) to 42 patients with inflammatory bowel disease on two occasions separated by 1 month. At the time the investigation was planned, the investigators made predictions about how change in the IBDQ score should relate to change in the other measures if this questionnaire was really measuring HRQL. Examples of the predictions and the results are as follows:

- The patient's global rating of change in disease activity should relate closely (correlation ~ 0.5) with change in the bowel-symptoms dimension of the Inflammatory Bowel Disease Questionnaire.
- **Correlation** observed was 0.42.
- Some relation (correlation ~ 0.3) should exist between change in the Disease Activity Index and change in the bowel-symptoms dimension of the Inflammatory

Bowel Disease Questionnaire.

- **Correlation** observed was 0.33.
- Some relation (correlation ~ 0.3) should exist between change in the Disease Activity Index and change in the systemic-symptoms dimension of the Inflammatory Bowel Disease Questionnaire.
- **Correlation** observed was 0.04.
- Change in the emotional-function dimension of the Inflammatory Bowel Disease Questionnaire should relate closely (correlation ~ 0.5) with change in the emotional-function dimension of the generic questionnaire.
- **Correlation** observed was 0.76.



Exercise 5: IBD Questionnaire

Did the results provide sufficient support for the validity of the questionnaire?



8. Interpretability

A final key property of an HRQL measure is *Interpretability*.

- For *discriminating between groups*, we could ask whether a particular score signifies that a patient is functioning normally or has mild, moderate, or severe impairment.
- For *evaluating change over time*, we might ask whether a particular change in a score represents a trivial, small but important, moderate, or large improvement or deterioration.
- A number of strategies are available to make PRO scores interpretable (Guyatt et al., 1991), such as:
 1. Classify patients into those who had important improvement as well as those who did not and examine the changes in score in the two groups;
 2. Interpret observed changes in PRO measures in terms of elements of those measures that will be familiar to readers (for instance, descriptions of changes in mobility); or
 3. Determine how scores in PRO measures relate to marker states that are familiar and meaningful to clinicians.

Data suggest that small, medium, and large effects correspond to changes of approximately 0.5, 1.0, and greater than 1.0 per question for instruments that present response options on seven-point scales (Jaeschke et al., 1989).



Example 5

- In a domain with 6 items, changes of 3 or 4 represent small effects, changes of 5 or 6 represent moderate effects, and changes of 7 or more represent large effects. Investigators used this information to interpret a recent trial that showed use of bronchodilators results in a small but clinically important improvement in dyspnea, fatigue, and emotional function in patients with chronic airflow limitation (Guyatt et al., 1987).
- In a study (Thompson et al., 1988) of patients with arthritis, a change of 0.02 points in the Quality of Well-Being utility instrument was equivalent to all treated patients improving from moving their own wheelchair without help to walking with physical limitations. The availability of data to improve the Interpretability of PROs is likely to increase exponentially in the next decade.

9. Types of Measures

There are a number of ways of categorizing instruments designed to measure PROs such as health status and quality of life (Fitzpatrick et al., 1992; Patrick and Erickson, 1993). Taxonomy of self-reported health status and quality of life measures is contained in Table 4.



Table 4: A Taxonomy of PRO Measures

MEASURE	STRENGTHS	WEAKNESSES
Source of Report		
Person and Patient	Sensations, feelings, evaluations known only to the person	When established that person cannot speak, write, or communicate to others
Proxy (Not a PRO)	Can observe and report behaviors only to patient	Cannot report feelings known
Mode of Collection		
Self-administered, with or without supervision	Privacy	Missing data, particularly by mail, and without supervision no assurance of who completed
Interviewer-administered	Control	Sensitive information like income sometimes difficult/cost
Computer-administered and/or computer-adapted	Flexibility	Cost/for persons not familiar or afraid of computer
Testing Strategy		
Dynamic or tailored content: based on health status, age, etc.	<ul style="list-style-type: none"> • Items relevant to person • More precise measurement 	Requires item bank and item calibration
Standard or fixed-length content	<ul style="list-style-type: none"> • Content same for all respondents • Easy to administer 	<ul style="list-style-type: none"> • Many content items not relevant to individual • Floor and ceiling effects

Types of Scores Produced		
Single indicator number	<ul style="list-style-type: none"> • Global evaluation • Sometimes easy to interpret 	May be difficult to interpret trends
Single index number	<ul style="list-style-type: none"> • Represents net impact • Useful for cost effectiveness 	Sometimes not possible to disaggregate contribution of domains to the overall score
Profile of interrelated scores	<ul style="list-style-type: none"> • Single instrument • Contribution of domains to overall score possible 	<ul style="list-style-type: none"> • Length may be a problem • May not have overall score
Battery of independent scores	Wide range of relevant outcomes possible	<ul style="list-style-type: none"> • Cannot relate different outcomes to common measurement scale • May need to adjust for multiple comparisons • May need to identify major outcome
Range of Populations and Concepts		
Generic: applied across diseases, conditions, populations, and concepts	<ul style="list-style-type: none"> • Broadly applicable • Summarizes range of concepts • Detection of unanticipated effects possible 	<ul style="list-style-type: none"> • May not be responsive to change • May not have focus of patient interest • Length may be a problem • Effects may be difficult to interpret
Specific: applied to individuals, diseases, conditions, populations, or concepts/domains	<ul style="list-style-type: none"> • More acceptable to respondents • May be more responsive to change 	<ul style="list-style-type: none"> • Cannot easily compare across conditions or populations • Cannot detect unanticipated effects

<p>Specific: applied to individuals, diseases, conditions, populations, or concepts/domains</p>	<ul style="list-style-type: none"> • More acceptable to respondents • May be more responsive to change 	<ul style="list-style-type: none"> • Cannot easily compare across conditions or populations • Cannot detect unanticipated effects
<p>Weighting System</p>		
<p>Utility: preference weights from patients, providers, or community</p>	<ul style="list-style-type: none"> • Interval scale • Patient or consumer view incorporated 	<ul style="list-style-type: none"> • May have difficulty obtaining weights • May not differ from statistical weights that are easier to obtain
<p>Equal-weighting: items weighted equally or from frequency or responses</p>	<ul style="list-style-type: none"> • Self-weighting samples • More familiar techniques • Appears easier to use 	<ul style="list-style-type: none"> • May be influenced by prevalence • Cannot incorporate tradeoffs

Generic instruments measure the full range of health and quality of life, without focusing on specific areas. They are designed for use across a wide variety of populations.

Specific instruments are designed for application to individuals, conditions or diseases, domains, or populations.

Generic and specific instruments may be *health profiles* or *utility measures*, which are distinguished by having preference weights applied to the items and domains. Some utility measures, and indeed some profiles, also incorporate an index score or single number for analyses. Utility measures are useful for economic applications, since they produce *quality-adjusted life years*, a combined measure of how long one lives as measured by survival or mortality and how well one lives, as measured by functional status and well-being.

10. Applications

Decision makers and analysts wanting to measure PROs should first identify the problem or application of the measure. With this information, one can then identify the desired characteristics of existing measures to be included in the assessment.

For example, monitoring the health of populations and communities demands parsimonious instruments including global evaluations across a number of conditions and different population groups. For comprehensive evaluation in a clinical trial, health profiles or batteries are most appropriate according to the main effects intended and unintended or adverse consequences of treatment. For economic evaluation, utility measures are useful to produce a comparison across alternative treatments.

PROs from Children

PROs directly from children and youth are taking greater prominence among all interested parties, following similar development as that for adults and older adults (Starfield, 1996). Children and youth represent a special case, however, given knowledge of variation in how and when children develop, the wide variation in willingness and ability to self-report across the age spectrum, and the “special” language of children and adolescents in different cultures of the world. Rapid progress is being made in Europe and North America, to be followed and informed by work in other parts of the world, often less accessible to Western parties.

Cross Cultural PROs

Demand is also increasing for quality of life measures available for use in cross-cultural comparisons, which requires special attention to cultural adaptation and validation for each culture in which the measure is applied (Acquadro et al., 2003).

The most desirable means of development and validation is to have the goal of cross-cultural comparability in mind from the beginning.

Measures developed simultaneously in different cultures have the advantage of identifying as early as possible those domains and items that are more or less valid in a particular culture or

population. Translating instruments developed in one culture for use in another is more common, but the danger of this approach is the assumption that the conceptual structure, domains, and items are cross-cultural.

Recognizing Cultural Differences






Assessments of functional status that use examples such as “ability to walk several blocks” run the danger that “blocks” have different if any meaning at all in different cultures. Response scale anchors, such as “quite a bit”, also do not translate easily into different languages.

The psychometric properties of cross-cultural measures is similar to that for instruments used within one language or cultural group, although standards for aggregation across sites have not been rigorously applied. When it is and is not valid to use measures in different populations and to pool data across different cultures remains an area for further investigation and debate.

11. Summary

Quality of life measurements are important for measuring the impact of disease, treatment, health and social policies, and the progress of economic and social development. Developers and users should specify and label the content and type of measure for every application of a PRO and provide evidence of its appropriateness to the intended use, for validity of the measure as used in a particular case, and how to interpret results. A major challenge faces developers and users of these measures in establishing a testable theory of the expected and observed relationships among the different concepts and domains of quality of life. Assessing the perspective of the person or patient is an important enterprise, and progress will be made most rapidly when we recognize that, as the poem notes, the measurement of 'Spring' requires careful description as to its purpose, its limits, and especially the evidence provided of where the content came from, what health or quality of life actually represents, and how the measures perform in application.

12. Resources

	<p>gateway.nlm.nih.gov/gw/cmd/</p> <p><i>MEDLINE, MEDLINEplus, OLDMEDLINE, PubMed, LOCATORplus, DIRLINE, AIDS Meetings, Health Services Research Meetings, Space Life Sciences Meetings, and Health Services Research Projects.</i></p>
	<p>www.golid.org</p> <p><i>Through the structured presentation of synthesized, reliable and constantly updated data on PRO instruments, the PROQOLID database aims to: Provide an overview of existing PRO instruments; Provide relevant and updated information on PRO instruments; Facilitate access to the instruments and their developers; Facilitate the choice of an appropriate PRO instrument. Descriptions of 436 instruments. (\$\$)</i></p>
	<p>www.olga-gol.com</p> <p><i>The most comprehensive source of information about questionnaires, rating scales and other tools for assessing psychosocial effectiveness in clinical and pharmacoeconomic studies. (\$\$)</i></p>
	<p>www.ovid.com</p> <p>Health and Psychosocial Instruments (HaPI)</p> <p><i>Medicine, Nursing, Health Professions, Social Sciences, Education; Over 100,000 surveys, questionnaires, scales, tests, checklists, manuals; CD-ROM or online subscription (\$\$); Abstracts from leading journals.</i></p>
	<p>www.qualitymetric.com</p> <p><i>Easy, valid, and reliable patient-reported outcome (PRO) measurement.</i></p> <p><i>Whether you require expertise in protocol/study design, information on the most cost-effective method of administering, scoring, interpreting and reporting on PRO Surveys, or simply want to license one of our industry-leading PRO surveys, QualityMetric makes PRO measurement easy and insightful, supporting your efforts to positively impact healthcare. (\$)</i></p>

	<p>www.measurementexperts.org</p> <p>Measurement Excellence and Training Resource Information Center (METRIC).</p> <p><i>Empowering Researchers with Measurement Knowledge. You will find resources to help you with your measurement needs. Find an instrument or learn about the Foundations of Measurement. METRIC is designed to give researchers the resources to measure their best.</i></p>
	<p>www.wiley.com/legacy/products/subject/reference/salek_toc.html</p> <p>The Compendium of Quality of Life Instruments (5-Volume Set plus CD-ROM) (\$\$)</p> <p>Compiled by: Sam Salek</p> <p><i>The first ever comprehensive source of original questionnaires and related information for researchers and practitioners in the field of health-related quality of life ...</i></p> <p><i>Reliable one-stop information plus CD-ROM, An all-in-one up-to-date sourcebook that is easy and quick to use, Tried and tested existing Quality of Life Instruments from around the world.</i></p>
	<p>www.atsqol.org</p> <p>American Thoracic Society – Quality of Life Resource</p> <p><i>The goal of this website is to provide information about quality of life and functional status instruments that have been used in assessing patients with pulmonary disease or critical illness.</i></p>
	<p>phi.uhce.ox.ac.uk</p> <p>Patient-reported Health Instruments Group (PHIG) National Centre for Health Outcomes Development (NCHOD)</p> <p><i>General information on patient-reported health outcomes and instrument selection, including guidance regarding the selection of appropriate instruments for use in clinical trials, systematic reviews of outcome measures relevant to specific disease and population (demographic) groups.</i></p>

*(\$\$) indicates fees involved for some

Guide for Reviewing PRO literature

This handout presents selected issues discussed in this Chapter that are specific to PROs that readers might consider when reading the research literature. Select a recent clinical trial from the literature. Apply the criteria in the checklist. Remember that there are no right or wrong answers to many of the questions. Few articles will meet all the criteria listed, and this checklist is simply a guide for reviewing the literature. Based on Chapter 7 of *Health Status and Health Policy*, Guyatt et al., 2008; Users' Guides to the Medical Literature: XII. How to Use Articles About Health-related quality of life; MOT measurement criteria and CDC Guide to Evaluation of Community Preventive Services.

13. References

- Acquadro C., Berzon R., Dubois D., Leidy N.K., Marquis P., Revicki D., Rothman M.; PRO Harmonization Group. (2003). Incorporating the patient's perspective into drug development and communication: An ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health*, Sep-Oct, 6(5), 522-31.
- Calman, K.C. (1987). Definitions and dimensions of quality of life. In Aaronson, N.K. & Beckman, J. (Eds.). *The Quality of Life of Cancer Patients*. Monograph series of the European Organization for Research on the Treatment of Cancer (EORTC). Vol 17. New York: Raven.
- Campbell A., Converse S.E., & Rodgers W.L. (1976). *The Quality of American Life*. New York : Russell Sage.
- Cook K.F., O'Malley K.J., Roddey T.S. (2005). Dynamic assessment of health outcomes: time to let the CAT out of the bag? *Health Serv Res*, 40(5 Pt 2), 1694-711.
- Doward, Lynda C. & McKenna, Stephen P. (2004). Defining Patient-Reported Outcomes. *Value in Health*, 7 (1 suppl), S4-S8.
- Fitzpatrick R., Fletcher A., Gore S., Jones D., Spiegelhalter D., Cox D. (1992). Quality of life measures in health care. I: application and issues in assessment. *British Medical Journal*, 305, 1074-7.
- Goodwin P.J., Black J.T., Bordeleau L.J., Ganz P.A. (2003). Health-related quality-of-life measurement in randomized clinical trials in breast cancer--taking stock. *Journal National Cancer Institute*, 95(4), 263-81.
- Guyatt G.H., Feeney D.H., Patrick D.L. (1991) Proceedings of the International Conference on the Measurement of Quality of Life as an Outcome in Clinical Trials: Postscript. June 14-17, 1989. *Controlled Clin Trials*, 12, 2665-269S.
- Guyatt G.H., Feeney D.H., Patrick D.L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine*, 118, 622-629.

Guyatt G.H., Thompson P.J., Berman L.B., Sullivan M.J., Townsend M., Jones N.L., Pugsley S.O.(1985). How should we measure function in patients with chronic heart and lung disease? *J Chronic Dis*, 38(6), 517-24.

Guyatt G.H., Townsend M., Pugsley S.O., Keller J.L., Short HD, Taylor DW, Newhouse MT. (1987). Bronchodilators in chronic airflow limitation. Effects on airway function, exercise capacity and quality of life. *Am Rev Respir Dis*, 135 (5), 1069-74.

Hornquist J.O.(1982). The concept of quality of life. *Scand J Soc Med*, 10(2), 57-61.

Jaeschke R., Singer J., Guyatt G. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Controlled Clin Trials*, 10(4), 407.15.

Joyce C.R.B., O'Boyle C.A., McGee H.M. (Eds). (1999). *Individual Quality of Life: Approaches to Conceptualization and Assessment*. London: Harwood Press.

Joyce C.R.B. (1988). Quality of Life: the state of the art in clinical assessment. In: Walker S.W., Rosser R.M. (Eds), *QoL: Assessment and Application*. (pp.169-179) Lancaster: MPT Press.

Kahneman D., Diener E., Schwarz N. (Eds). (1999). *Well-being: The Foundations of Hedonic Psychology*. New York : Russell Sage.

Kunz R., Bucher H.C., Cook D., Holbrook A., Guyatt G. (2008) *Users' Guide to the Medical Literature*. In: G. Guyatt, D. Rennie, M. Meade, D. Cook. (Eds). New York: McGraw-Hill.

Lai J.S., Cella D., Chang C.H., Bode R.K., Heinemann A.W. (2003). Item banking to improve, shorten and computerize self-reported fatigue: An illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Qual Life Res*, 12(5), 485-501.

Maslow, A. (1943). A theory of human motivation. *Psychol Rev*, 50, 370-96.

McHorney C.A., Cohen A.S. (2000). Equating health status measures with item response theory: illustrations with functional status items. *Med Care*, 38(9 suppl), I143-59.

McKenna S.P., Doward L.C., Meads D. Patrick D.L. & Tennant A. (2004). Summary of Needs-Based Quality of Life Instruments. *Value in Health*, 7(s1), S39-S40.

McKenna S.P., Doward L.C., Niero M., Erdman R. (2004). Development of needs-based quality of life instruments. *Value Health*, 7(1 suppl), S17-S21.

McKenna, Stephen P. & Doward, Lynda C. (2004). The needs-based approach to quality of life assessment. *Value in Health*, 7(1 suppl), S1-S3.

Patel K.K., Veenstra D.L., Patrick D.L. (2003). A review of selected patient-generated outcome measures and their application in clinical trials. *Value in Health*, 6(5), 595-603.

Patrick D.L., Bergner M. (1990). Measurement of health status in the 1990s. *Annual Rev Public Health*, 11,165-183.

Patrick D.L., Burke L.B., Powers J.H., Scott J.A., Rock E.P., Dawisha S., O'Neill R., Kennedy, D.L. (2007). Patient-reported outcomes to support medical product labeling claims. *Value in Health*, Nov; 10 (s2 The FDA Guidance for Patient-Reported Outcomes), S125-S137.

Patrick D.L., Chiang Y.P. (2000). Measurement of health outcomes in treatment effectiveness evaluations: Conceptual and methodological challenges. *Medical Care*, 38(9 suppl), I114-25.

Patrick D.L., Danis M., Southerland L.I., Hong G. (1988). Quality of life following intensive care. *J Gen Intern Med*, 3(3), 218-23.

Patrick D.L., Deyo R.A. (1989). Generic and disease-specific measures in assessing health status and quality and life. *Medical Care*, 27 (3 suppl), S217-S232.

Patrick D.L., Erickson P. (1993). *Health status and health policy: Quality of life in health care evaluation and resource allocation*. New York: Oxford University Press.

Patrick D.L., Martin M., Bushnell D., Pesa J. (2003). Measuring satisfaction with migraine treatment: expectations, importance, outcomes, and global rating. *Clinical Therapeutics*, 25 (11), 2920-2935.

Patrick D.L. (2003). Patient-Reported Outcomes (PROs): An organizing tool for concepts, measures, and applications. *Quality of Life Newsletter*, 31, 1-5.

Rogerson R.J. (1995). Environmental and health-related quality of life: conceptual and methodological similarities. *Social Science and Medicine*, 41(10), 1373-1382.

Scientific Advisory Committee, Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, 11(3), 193-205.

Skevington S.M., O'Connell K.A.; WHOQOL Group. (2004). Can we identify the poorest quality of life? Assessing the importance of quality of life using the WHOQOL-100. *Qual Life Res*, 13(1), 23-34.

Starfield B. (1996). Health status measurement: the special case of children and youth [editorial]. *Inj Prev*, 2(2), 86-7.

Streiner D.L., Norman G.R. (1995). *Health measurement scales: A practical guide to their development and use*. New York: Oxford University Press.

Szabo, S. (1996). On behalf of the WHOQOL Group The world health organization quality of life (WHOQOL) assessment instrument. In Spiker, B. *Quality of Life and Pharmacoeconomics in Clinical Trails*. (2Ed.) Philadelphia: Lippincott-Raven Publishers.

Thompson M.S., Read J.L., Hutchings H.C., Paterson M., Harris E.D. Jr. (1988). The cost effectiveness of auranofin: results of a randomized clinical trial. *J Rheumatol*, 15(1), 35-42.

Ware J.E. Jr, Kosinski M., Bjorner J.B., Bayliss M.S., Batenhorst A., Dahlof C.G., Tepper S., Dowson A. (2003). Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res*, 12(8), 935-52.

WHOQOL Group. (1993). WHOQOL Group, Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL), *Quality of Life Research* 2, 153–159.

WHOQOL Group. (1994). WHOQOL Group, The Development of the World Health Organization Quality of Life Assessment Instrument. In: J. Orley and W. Kuyken, (Eds.) *Quality of life assessment: International perspectives*, Springer, Berlin (1994), pp. 41–57.

WHOQOL Group. (1995). WHOQOL Group, The World Health Organization Quality of Life Assessment: Position Paper from the World Health Organization, *Social Science and Medicine* 41, 1403–1409.

WHOQOL Group (1996). WHOQOL Group, What quality of life? The World Health Organization Quality of Life Assessment, *World Health Forum* **17**, 354–356.

WHOQOL Group. (1998). WHOQOL Group, The World Health Organization Quality of Life Assessment (WHOQOL): Development and general psychometric properties. *Social Science and Medicine* **46**, 1585–1596.

Wilson I.B., Cleary P.D. (1995). Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. *JAMA*, 273(1), 59-65.

14. Author Biographies

Donald Patrick, PhD was founding president of the International Society for Quality of Life Research and is co-convenor of the Quality of Life Methods Group of the Cochrane Collaboration. He works on the assessment of health status and quality of life in application to adolescents, chronic illness and disability, health promotion, evaluation research, underserved populations, and dying and death. He was an original developer of the generic Quality of Well-Being and Perceived Quality of Life Scales, and has developed numerous cross-cultural instruments. He directs the Seattle Quality of Life Group, co-directs the End of Life Care Research Group and teaches the assessment of health outcomes at the University of Washington. He is a special government employee for the US Food and Drug Administration in developing guidance and training for the use of patient-reported outcomes in the evaluation of medical products. He consults widely throughout the world on outcomes assessment.

Gordon Guyatt, MD is a physician and Professor of Clinical Epidemiology & Biostatistics at McMaster University in Hamilton, Ontario. He is known for his work on evidence-based medicine, a term that first appeared in a paper he published. He has published over 600 peer-reviewed articles in medical journals. He has also written extensively on health care policy in the popular press. He is widely known for his teaching, mentoring and advocacy. His contributions to quality of life research, randomized trials and meta-analysis have been considered groundbreaking. In 2007, the BMJ launched an international election for the most important contributions to healthcare. Evidence-based medicine came 7th in the line-up, beating the computer. In 1979, Guyatt and Dr. Fred Freedman co-founded the Medical Reform Group, a Canadian organization of physicians, medical students and others which supports universal public health care. Guyatt continues to serve as a member of the steering committee and leading spokesperson for this group.