Validity of Quality of Life Measurement Tools — From Generic to Disease-specific

GEORGE A. WELLS, ANTHONY S. RUSSELL, BOULOS HARAOUI, ROBERT BISSONNETTE, and CARL F. WARE

ABSTRACT. Health-related quality of life (HRQOL) is an important measure of a patient's perception of his/her illness. Over the past 3 decades, numerous instruments have been developed to measure HRQOL in various patient populations, with 2 basic approaches: generic and disease-specific. While generic measures have broad application across different types and severity of diseases, disease-specific measures are designed to assess particular diseases or patient populations. All HRQOL instruments, however, must be valid and have high reliability and responsiveness. Validity ensures that the instrument measures what it is supposed to measure. Reliable instruments are able to reproducibly differentiate between subjects. Responsive evaluative measures are able to detect important changes in HRQOL during a period of time, even if those changes are small. HRQOL measures should also be interpretable, meaning that the differences in scores that correspond to small, moderate, and large HRQOL changes are easily identifiable. This article describes the steps in the development of HRQOL instruments from the conceptual framework to creation and testing. Several examples of generic and disease-specific instruments commonly used to evaluate HRQOL in patients with immune-mediated inflammatory diseases (IMID) are provided. (J Rheumatol 2011;38 Suppl 88:2–6; doi:10.3899/jrheum.110906)

Key Indexing Terms:
HEALTH-RELATED QUALITY OF LIFE
GENERIC MEASUREMENT INSTRUMENTS
RESPONSIVENESS
RELIABILITY
DISEASE-SPECIFIC

From the University of Ottawa Heart Institute, Ottawa, Ontario; University of Alberta Hospital, Edmonton, Alberta; Centre Hospitalier de l'Université de Montréal (CHUM) Hôpital Notre-Dame; Innovaderm Research, Montreal, Quebec, Canada; and Sanford Burnham Medical Research Institute, La Jolla, California, USA.

Supported by an unrestricted grant from Abbott Canada. Dr. Wells has received funds from BMS, Abbott, Amgen, and UCB. Dr. Russell has received consultant fees from Amgen/Wyeth, Schering-Plough, UCB, and Abbott Canada, and speaking fees from Roche and BMS. Dr. Haraoui has acted as consultant for Abbott Laboratories, Amgen, Bristol-Myers Squibb Canada, Roche Canada, Schering-Plough Canada, UCB Pharma, and Wyeth Pharmaceuticals; received grant/research support from Abbott Laboratories, Amgen, Bristol-Myers Squibb Canada, Roche Canada, Schering-Plough Canada, UCB Pharma, and Wyeth Pharmaceuticals; and received honoraria from Abbott Laboratories, Amgen, Bristol-Myers Squibb Canada, Roche Canada, Schering-Plough Canada, UCB Pharma, and Wyeth Pharmaceuticals. Dr. Bissonnette received research grants or honoraria from Abbott, Amgen, Astellas, Merck, Pfizer, and Schering.

G.A. Wells, PhD, Professor, Department of Epidemiology and Community Medicine, University of Ottawa, Professor, Department of Medicine and Senior Scientist, University of Ottawa Heart Institute, Director, Cardiovascular Research Methods Centre, University of Ottawa Heart Institute; A.S. Russell, MA, MB, BChir, FRCPC, Professor Emeritus, Department of Medicine, University of Alberta; B. Haraoui, MD, FRCPC, Associate Professor, Université de Montréal, and Director of Clinical Research, Department of Rheumatology, Centre Hospitalier de l'Université de Montréal (CHUM) Hôpital Notre-Dame; R. Bissonnette, MD, FRCPC, Medical Director, Innovaderm Research; C.F. Ware, PhD, Director, Infectious and Inflammatory Diseases Center, and Professor, Laboratory of Molecular Immunology, Sanford Burnham Medical Research Institute.

Address correspondence to Dr. G. Wells, Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Room H1281, 40 Ruskin Street, Ottawa, Ontario K1Y 4W7.
E-mail: gawells@ottawaheart.ca

Health-related quality of life (HRQOL) is a multidimensional concept referring to patients' perceptions of the influence of disease and treatment on their physical, psychological, and social function and well-being[1]. To that end, various HRQOL instruments have been developed[2]. HRQOL questionnaires can be self- or interviewer-administered. They are used to measure either differences in quality of life between patients at any point in time or changes in HRQOL within patients during a certain period of time[3].

HRQOL instruments are valuable for a number of purposes, including evaluating the impact of disease and/or treatment on a patient's overall well-being in clinical trials and daily practice[4]. HRQOL tools can also be used for assessing health-related gaps across different segments of the population and for measuring and comparing the effectiveness of healthcare interventions for various conditions[5]. Further, both clinicians and policymakers are recognizing the importance of measuring HRQOL in the daily management of patients with various diseases, as well as in making policy-related decisions[3]. In the current environment with drugs and biologics that cost more than older compounds, improvements in HRQOL are usually a requirement for coverage of these medications. HRQOL instruments can be important tools to confirm that highly statistically significant improvement in a disease is associated with improvement in HRQOL.

INSTRUMENT DEVELOPMENT

The process of the development of a HRQOL instrument entails a series of steps, including:

1. Identification of concepts and development of a conceptual framework
2. Instrument creation
3. Assessment of instrument properties, and
4. Instrument modification[6].

Each step requires a different study construct. While the initial phases seek to establish the content and the format of the instrument, the later phases aim to determine the measurement properties of the instrument.

Identification of concepts and development of conceptual framework. The initial, and probably the most important, step in instrument development is to identify concepts and domains of importance[7]. The most commonly considered domains include psychological, physical, and social functioning as well as somatic comfort[7,8,9]. When deciding on these, one must consider the intended population (i.e., adults, children), condition, timeframe, and research application. Expected relationships among concepts are to be hypothesized.

Instrument creation. During this phase, a first draft of the instrument is generated. Table 1 outlines a list of items and considerations included in the process.

Items for inclusion in the measure are usually derived from a literature review, accompanied by a series of interviews with patients and experts in the field. Administrative methods, response period, and response scales are chosen. It is important to keep a manageable number of items by identifying and choosing those that are (1) most pertinent to the purpose of the measure, (2) most important to the patient, and (3) most frequent. The selection of the number of items is generally performed by a panel of experts in a particular therapeutic field. Thereafter, the instrument is formatted and the instructions are drafted. It is important to ensure that the items and their description make sense, are easily understood, and are easy to complete within an allocated timeframe. The types of scoring of an instrument can vary, namely: a single rating on a single concept; a single score combining multiple ratings of related domains or individual concepts (index); multiple uncombined scores of multiple related domains (profile); multiple uncombined scores of independent concepts (battery); or a combination of single rating, index, profile, or battery (composite). Once the initial version is developed, it must undergo pilot testing, after which the instrument and the procedures are refined.

Table 1. Items considered during the initial development of a health-related quality of life (HRQoL) instrument.

Number of items	Single item for single concept
	Multiple items for a single concept
	Multiple items for multiple domains within a concept
Intended measurement	Generic vs specific (condition-specific, population-specific)
Data collection method	Interviewer-administered
	Self-administered
	Interactively administered (computer-assisted, web-based interactive voice response)
Timing	As events occur; at defined intervals throughout a study
	Timeframe: within the last "period" (i.e., within last week, 2 weeks, month, etc.)
Types of scores	Single rating on a single concept
	Index (single score combining multiple ratings of related domains or individual concepts
	Profile (multiple uncombined scores of multiple related domains)
	Battery (multiple uncombined scores of independent concepts)
	Composite (single rating, index, profile, or battery)
Weighting of items or concepts	All items and domains are equally weighted
	Items are assigned variable weights
	Domains are assigned variable weights
Response formats	Visual analog scale, Likert scale, rating scale, checklist

Assessment of instrument properties. The third phase in the development of an HRQOL instrument includes testing to ensure its reliability, validity, and ability to detect change. The reliability of an instrument to consistently measure the characteristic of interest includes the concepts of both reproducibility (rater agreement) and internal consistency (correlation among questions composing an instrument so that different questions on the same concept are in agreement). Administrative and respondent burden is evaluated and items are revised accordingly. After the identification of meaningful differences in scores, the format, scoring procedures, and training material are finalized.

Instrument modification. Finally, during the modification phase, the concepts measured, populations studied, research application, and instrumentation or method of administration can be modified further, if necessary.

EVALUATION OF AN INSTRUMENT AND ITS PROPERTIES

Once the initial version of an instrument is developed, it must be tested to ensure that it is reliable, valid, and responsive to changes.

The Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) filter. OMERACT is an international organization that focuses on rheumatology outcome measures[10]. To be accepted as an OMERACT-endorsed outcome measure for use in a clinical trial, the instrument must pass through the OMERACT filter in its intended setting[10]. The filter has 3 component criteria, each represented by a question to be answered about that measure: truth, discrimination, and feasibility.

In addition, the following 5 criteria should be applied (Table 2): (1) *Feasibility* ensures that the instrument is used efficiently. It takes into consideration administration time, reading and understanding level, and multicenter administration.

Table 2. Criteria for QoL instruments.

Feasibility	Can the instrument be used efficiently?
Reliability	Does the instrument consistently measure the characteristic of interest? (Includes both the concepts of rater reproducibility and internal consistency of questions)
Validity	Does the instrument measure what it is supposed to measure?
Responsiveness	Can the instrument detect important changes?
Interpretability	Has the clinical importance of change been established?

(2) *Reliability* certifies that the instrument consistently measures the characteristics of interest. It includes reproducibility and internal consistency. Repeated testings are performed to verify test-retest, intrarater, and interrater reliability. In summary, a reliability value of 0.00 indicates absence of reliability and value of 1.00 means perfect reliability. A reliability coefficient > 0.70 is desirable as it implies that 70% of the measured variances are reliable and 30% are owed to random error[11]. Cohen's κ coefficient is a statistical measure of interrater agreement and values > 0.60 are desirable as they show good to excellent agreement between the 2 raters' scores[12].

Internal consistency represents the correlations among items composing an instrument. It verifies whether several items that propose to measure different aspects of the same trait produce similar scores. The goal of a reliable instrument is for scores on similar items to be related, but for each to contribute some unique information. Internal consistency is usually measured with Cronbach's α, which ranges between 0 and 1[13]. A rule of thumb is that $\alpha = 0.7$ indicates acceptable, and $\alpha = 0.8$ represents good reliability. However, reliabilities of $\alpha = 0.9$ are not necessarily desirable, as this indicates that the items may be redundant.

(3) *Validity* confirms that the instrument measures what it is supposed to by demonstrating appropriate correlations with other measures, usually based on a prior prediction of the degree of such correlations. Validity of content, criterion, and construct are the standard categories.

(4) *Responsiveness* verifies the ability of an instrument to detect small but clinically important change. This is particularly important where subjective reports of health status are one of the primary outcomes of the trial. There are several ways to present responsiveness: difference from baseline, treatment difference, relative percentage improvement, mean response, and relative efficiency. Receiver operator characteristic curves (Figure 1)[14] evaluate how a given change in score can discriminate patients who improve from those who do not.

(5) *Interpretability* ensures that the minimal clinically important difference (MCID) is established. Table 3 provides an overview of established MCID for 2 commonly used HRQOL instruments: the Medical Outcomes Study Short-Form Survey (SF-36) and EuroQoL (EQ-5D)[15,16,17,18,19,20].

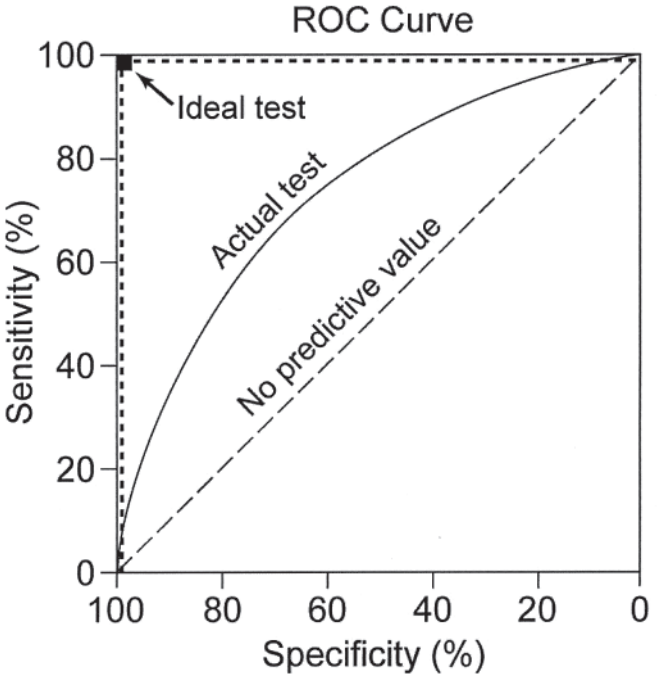

Figure 1. Example of a receiver operator characteristic (ROC) curve. Used with permission from Sprawls P. The physical principles of medical imaging. 2nd ed.[14] [Internet. Accessed August 9, 2011.] Available from: http://www.sprawls.org/ppmi2/IMGCHAR

Table 3. The minimal clinically important difference (MCID) for Medical Outcome Study Short-Form 36 (SF-36) and EuroQoL (EQ-5D).

Instrument	MCID
Short Form	
SF-36 domains[15]	3 to 5
SF-36 / SF-12	PCS: 1.6 to 7.0
PCS/MCS[16]	MCS: 2.3 to 8.7
SF-6D[17,18]	0.03 to 0.041
EuroQol EQ-5D	
EQ-5D[19]	Estimated 0.033 to 0.074
EQ-5D VAS[16]	4.2 to 14.8
EQ-5D utility[20]	0.05

PCS/MCS: physical and mental component summaries; VAS: visual analog scale

GENERIC VERSUS DISEASE-SPECIFIC INSTRUMENTS

HRQOL measures can be divided into generic and disease-specific[21].

Generic instruments. Generic instruments are designed to assess HRQOL in a broad range of populations with or without chronic illness. While generic instruments may not be sufficiently sensitive to detect changes in HRQOL in any specific illness, they are used to compare HRQOL in popu-

lations with different diseases. Further, when these measures are applied to healthy populations, normative data can be gathered. These data can then be used to compare the burden of disease of a specific condition to that of other chronic illnesses, as well as to healthy controls. Commonly used generic measures include the SF-36[22] and EQ-5D[23].

The SF-36 is a short-form measure of generic health status in the general population (Table 4). It is designed for self-administration and it can be administered to anyone over the age of 14 years. It consists of 36 items divided into 8 health profiles. The SF-36 has been translated and adapted in many countries.

EQ-5D consists of descriptions of health status according to 5 dimensions (Table 5)[24]. Each dimension is divided into 3 levels. By combining different levels from each dimension, EQ-5D defines a total of 243 health states. These states may be converted to a score using "sets of values" derived from general population samples. EQ-5D provides a simple descriptive profile and generates a single index value for heath status on which full health is assigned a value of 1 and death a value of 0.

Table 4. Medical Outcome Study Short-Form 36 health domains[15,22].

Domain	Description
Physical functioning	Limitations in physical activity because of health problems (10 items)
Social functioning	Limitations in social activities because of physical or emotional problems (2 items)
Role limitations — physical	Limitations in usual role activities because of physical health problems (4 items)
Bodily pain	Presence of pain and limitations due to pain (2 items)
General medical health	Self-evaluation of personal health (5 items)
Mental health	Psychological distress and well-being (5 items)
Role limitations — emotional	Limitations in usual role activities because of emotional problems (3 items)
Vitality	Energy and fatigue (4 items)

Table 5. EuroQoL: EQ-5D. Reprinted from EuroQol Group. Health Policy 1990;16:199-208, with permission from Elsevier.

Mobility	☐ I have no problems in walking about
	☐ I have some problems in walking about
	☐ I am confined to bed
Self-care	☐ I have no problems with self-care
	☐ I have some problems washing or dressing myself
	☐ I am unable to wash or dress myself
Usual activities	☐ I have no problems with performing my usual activities
	☐ I have some problems with performing my usual activities
	☐ I am unable to perform my usual activities
Pain/discomfort	☐ I have no pain or discomfort
	☐ I have moderate pain or discomfort
	☐ I have extreme pain or discomfort
Anxiety/depression	☐ I am not anxious or depressed
	☐ I am moderately anxious or depressed
	☐ I am extremely anxious or depressed

Disease-specific instruments. Disease-specific instruments focus on concerns relevant to a particular illness. These instruments measure changes in HRQOL over time or with treatment, which is not possible with generic measures. Some of the common disease-specific instruments used in IMID include: ASQoL for ankylosing spondylitis[25], PsAQoL for psoriatic arthritis[26], RAQoL for rheumatoid arthritis[27], PDI (Psoriasis Disability Index)[28], IBDQ (Inflammatory Bowel Disease Questionnaire)[29], and IBDQoL[30].

Specific instruments can also be considered in a broader context, namely system- or organ-specific instruments. For example, the most frequently used instrument to assess HRQOL in patients with psoriasis is the Dermatology Life Quality Index (DLQI). This is neither a generic (it is centered on skin) nor a disease-specific instrument. An advantage with this type of questionnaire is its application for assessment of other dermatological diseases; physicians not involved in research prefer it to using and interpreting 10 or 20 different disease-specific questionnaires. However, it also has disadvantages. For example, some skin diseases also involve other organs (psoriasis involves skin and joints) and the DLQI does not assess the influence of the joint component on QOL.

CONCLUSION

HRQOL instruments allow broader assessment of the effects of disease and intervention on patients. They are applicable to all phases of trial assessment and require active patient participation. They also provide a standardized tool for comparison with other studies. The incorporation of HRQOL instruments in a study improves the likelihood of uptake by decision-makers and healthcare providers. Results obtained from QOL assessments are also considered valuable and are often used in pharmacoeconomic evaluations.

REFERENCES

1. Fihn SD. The quest to quantify quality. JAMA 2000;283:1740-2.
2. Patient-Reported Outcome and Quality of Life Instruments Database. Lyon, France: Mapi Research Trust. [Internet. Accessed Aug 15, 2011.] Available from: http://www.proqolid.org
3. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622-9.
4. Guyatt GH, Veldhuyzen Van Zanten SJ, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: a taxonomy and review. CMAJ 1989;140:1441-8.
5. The Health Measurement Research Group. Health-related quality of life measures. Madison, WI: University of Wisconsin. [Internet. Accessed Aug 15, 2011.] Available from: http://www.healthmeasurement.org/Measures.html
6. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985;38:27-36.
7. Cella DF, Tulsky DS. Measuring quality of life today: methodological aspects. Oncology (Williston Park) 1990;4:29-38.

8. Bonomi AE, Shikiar R, Legro MW. Quality-of-life assessment in acute, chronic, and cancer pain: A pharmacist's guide. J Am Pharm Assoc (Wash) 2000;40:402-16.
9. Schipper H, Levitt M. Measuring quality of life: risks and benefits. Cancer Treat Rep 1985;69:1115-25.
10. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: An international initiative to improve outcome measurement in rheumatology. Trials 2007:8:38.
11. Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika 1973;38:1-10.
12. Flack VF, Afifi AA, Lachenbruch PA, Schouten HJA. Sample size determinations for the two rater kappa statistic. Psychometrika 1988;53:321-5.
13. Bland JM. Altman DG. Statistics notes: Cronbach's alpha. Br Med J 1997;314:572.
14. Sprawls P. Image characteristics and quality. The physical principles of medical imaging. 2nd ed. Medical Physics Pub.; 1995. Montreat, NC: Sprawls Educational Publishing. [Internet. Accessed Aug 15, 2011.] Available from: http://www.sprawls.org/ppmi2/IMGCHAR
15. Hays RD, Morales LS. The RAND-36 measure of health-related quality of life. Ann Med 2001;33:350-7.
16. Coteur G, Feagan B, Keininger DL, Kosinski M. Evaluation of the meaningfulness of health-related quality of life improvements as assessed by the SF-36 and the EQ-5D VAS in patients with active Crohn's disease. Aliment Pharmacol Ther 2009;29:1032-41.
17. Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes 2003;1:4.
18. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 2005;14:1523-32.
19. Sullivan PW, Ghushchyan V. Preference-Based EQ-5D index scores for chronic conditions in the United States. Med Decis Making 2006;26:410-20.
20. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095-108.
21. Patrick L, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care 1989;27: S217-32.
22. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
23. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33:337-43.
24. EuroQol Group. EuroQol — a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199-208.
25. Doward LC, Spoorenberg A, Cook SA, Whalley D, Helliwell PS, Kay LJ, et al. Development of the ASQoL: a quality of life instrument specific to ankylosing spondylitis. Ann Rheum Dis 2003;62:20-6.
26. McKenna SP, Doward LC, Whalley D, Tennant A, Emery P, Veale DJ. Development of the PsAQoL: a quality of life instrument specific to psoriatic arthritis. Ann Rheum Dis 2004;63:162-9.
27. de Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. Br J Rheumatol 1997;36:878-83.
28. Finlay AY, Kelly SE. Psoriasis — an index of disability. Clin Exp Dermatol 1987;12:8-11.
29. Guyatt G, Mitchell A, Irvine EJ, Singer J, Williams N, Goodacre R, et al. A new measure of health status for clinical trials in inflammatory bowel disease. Gastroenterology 1989;96:804-10.
30. Patrick DL, Drossman DA, Frederick IO, DiCesare J, Puder KL. Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Dig Dis Sci 1998;43:400-11.