Taylor&Francis healthsciences

Development and Validation of QoL5 for Clinical Databases. A Short, Global and Generic Questionnaire Based on an Integrated Theory of the Quality of Life

Jes S. Lindholt[1], Søren Ventegodt[2] and Eskild W. Henneberg[3]

From the [1]Department of Vascular Surgery, Rigshospitalet, [2]Quality of Life Research Centre, Copenhagen, [3]Department of Vascular Surgery, Viborg Hospital, Denmark

Eur J Surg 2002; 168: 107–113

ABSTRACT

Objective: To develop and validate a short, global, and generic quality of life (QoL) questionnaire for clinical databases. The construct validity and item weighting of existing questionnaires are increasingly questioned.

Design: Cross-sectional population study.

Subject: 2460 Danes aged 18–88 years, randomly selected through the Danish Central Person Registry.

Interventions: Ten questions covering the spectrum of the integrative theory of QoL together with the Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), and self-estimated QoL questionnaire were sent by mail. A test-retest study of 50 people was conducted after one month.

Main outcome measures: Construct and criterion validity, reliability, and sensitivity.

Results: QoL5 correlations with SIP, NHP, Self-estimated QoL were 0.37, 0.52, and 0.76, respectively, and increased among those who were unwell. Cronbach's α was 0.69. All correlations in Siegel's test were over 0.6, and the test-retest correlation was 0.82. Only 12 respondents in each group will be needed to detect a difference of 10% in the QoL score between two groups.

Conclusions: QoL5 is a valid global and generic QoL measurement. Despite the use of only five questions, internal consistency and sensitivity were acceptable. So a relevant and practical outcome measurement is available for clinical databases.

Key words: quality assurance, global, generic, surgery, registry.

INTRODUCTION

During the past decades quality of life (QoL) has become a central outcome for treatment, prevention, and psychosocial support (6, 8, 21, 22). Clinical databases have been or are being developed in almost all specialities to record the activity for prognostic calculations, and to secure the quality of the treatments overall and in individual treatment units. Traditional outcomes such as mortality and morbidity have been included, but rarely QoL measurements (19). This may be because of practical difficulties concerning the implementation or difficulties in selecting the proper and most rational measurement. Most existing measurements were constructed without a solid philosophy of QoL, and were developed and validated ad hoc, using statistics to calculate the weighting of the items (1, 2, 4, 7, 14). Consequently, their weighting and whether they actually measure QoL are incomprehensible.

However, the concept of QoL and the good life have recently been subject to philosophical and psychological considerations, particularly in Scandinavia (1, 10, 11, 16, 19, 22). The integrative philosophy of QoL seems to build a bridge between the existing questionnaires and these considerations. It was presented first by Ventegodt in 1996 (19), and commented on by Fitzpatrick in BMJ later the same year (9). The theory consists of a continuous range of subjective QoL to objective QoL through QoL in the deep existential QoL, thus integrating most existing theories of QoL (Figure 1). Existential QoL refers to the state of *humanity's inner depth*—the inner state of a person's life or the state of the soul as explained by thinkers such as Kierkegaard, Sartre, Maslow, Frankl, and Antonovsky. Consequently, we have developed and validated a rational, short, generic and global QoL questionnaire based on this integrative theory of QoL for use in clinical databases.

SUBJECTS AND METHODS

Ten questions covering the integrative philosophy of QoL were formulated and used to define three short QoL measurements:

 ISSN 1102–4151

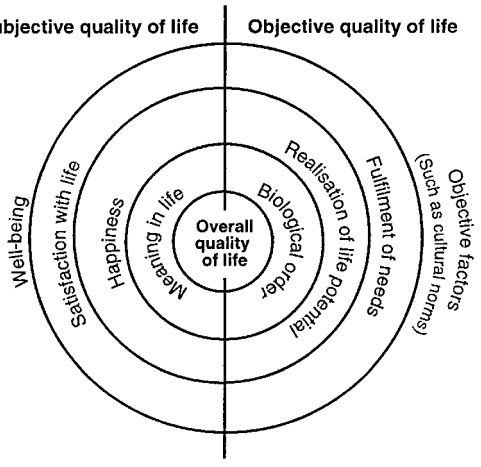

Fig. 1. The integrative theory of quality of life. A person can best be compared to a green apple with red patches (a subjective and an objective quality of life, respectively, at the surface of an individual's existence) with a hidden nucleus (humanity's inner depth). When this picture is combined with the pictures of humanity as an onion with a number of layers between the surface and the nucleus, the taxonomy underlying quality-of-life analysis is explained. Between life's surface and its inexpressible depth lie well-being, satisfaction, harmony, meaning and concord (19).

QoL1: simply asking about the QoL.
QoL5: 1, 2, and 2 questions about subjective, objective, and existential QoL, respectively.
QoL9: Three questions about subjective, objective, and existential QoL.

Example: QoL1: How would you assess the quality of your life now?
Answer: 1: very high, 2: high, 3: neither high nor low, 4: low, 5: very low.

Through interviews, philosophical considerations, and after revisions, the questions were refined to be unambiguous, independent, and different—that is, not obtaining the same information, and collectively covering the areas of the philosophy. Consequently, each question was given the same weight within each QoL measurement. Five-step Likert scales were used with a neutral point in the middle. This scale also contains a layout so that it also contains a subliminal visual analogue scale and a subliminal numerical scale, the combination of the three scales giving fair but not absolute meaning to the numerical scoring of "Likert answer 1 = very high" into 90%, "answer 2 = high" into 70%, "answer 3 = neither high nor low" into 50%, "answer 4 = low" into 30% and "answer 5 = very low" into 10% (19).

The mean of each of the three aspects of QoL was calculated, and the total QoL score was calculated as the mean of these three scores.

A questionnaire consisting of the ten questions, Nottingham Health Profile (NHP) (12, 13), Sickness Impact Profile (SIP) (3, 15), and self estimated QoL (19), was designed and sent to 2460 anonymous men and women aged 18–88, randomly selected from the Danish Central Register of Persons recording all Danes. Those who had contacted a doctor within the previous month and had been given drugs were classified as unwell (N = 1100).

Finally, to analyse reproducibility, a test-retest was conducted within a month on 50 patients who had been operated on for peripheral ischaemia of the lower limbs.

With respect to validity, the integrated QoL theory forms the basis of construct validity according to the methodological requirements for questionnaire-based QoL research (19). Criteria validity was calculating using with calculation of Spearman's correlation coefficient for NHP, SIP, and Self-estimated QoL.

With respect to reliability, reproducibility was tested with Spearman's correlation coefficients, and variance of the absolute and arithmetical differences, the latter as described by Bland and Altman (5). Internal consistency was analysed by calculation of Cronbach's α (14) with comparison with the three other measurements, and by correlation of each item with the QoL-scores as described by Siegel (17).

Finally, sensitivity was calculated to evaluate the number of respondents in both groups that were needed to detect a 3%, 10%, and 20% difference in QoL between the two groups. The questionnaire was approved by the relevant Danish committee of biomedical ethics. A total of 1100 respondents (44.7%) completed the questionnaire.

Table I. *Criteria validation of the new short measurements of QoL: QoL1, QoL5, and QoL9 by correlation analyses with the Nottingham Health Profile (NHP), the Sickness Impact Profile (SIP), and the Self-estimated quality of life questionnaire (SeQoL) among the whole population and the unwell subgroup*

	Whole population $n = 1100$			Unwell subgroup $n = 374$		
	SIP	NHP	SeQoL	SIP	NHP	SeQoL
QL1	0.35	0.48	0.68	0.40	0.52	0.66
QL5	0.38	0.52	0.76	0.47	0.59	0.73
QL9	0.44	0.58	0.80	0.54	0.68	0.77

RESULTS

Validity

The results of the validation of the criteria are shown in

Table II. *Internal consistency. Numerical values of Cronbach's α concerning raw and standardised variables*

	Whole population $n = 1100$		Unwell subgroup $n = 374$	
	Raw	Standardised	Raw	Standardised
QoL1	1.00	1.00	1.00	1.00
QoL5	0.74	0.74	0.74	0.73
QoL9	0.78	0.78	0.76	0.76
Self-estimated QoL	0.71	0.75	0.69	0.73
Nottingham Health Profile	0.76	0.78	0.76	0.77
Sickness Impact Profile	0.80	0.85	0.80	0.85

Table I. The correlations with NHP, SIP, and Self-estimated QoL are shown for all respondents and for these who were unwell. Apart from the correlations with SIP, all correlations concerning QoL5 and QoL9 were approximately 0.5 and 0.6 or above, respectively.

Reliability

Internal consistency was evaluated by Cronbach's α (14) and Siegel's test (17).

In Table II, Cronbach's α concerning total score and the SDs are compared between the new short measurements and the chosen reference standards. The calculation of Cronbach's α includes the number of items in the measurement, so favours questionnaires with many items. However, the α of QoL5 and QoL9 were similar to those chosen as reference standards, but slightly weaker for the values of NHP and SIP among those who were unwell.

The correlations between the individual items and the QoL scores ("Siegel's test") are listed in Table III. Apart from one item in QoL9, all items had correlations of 0.58 or more to their aspect of QoL and the total QoL-score.

The results concerning the test-retest showing the reproducibility are shown in Table IV. Seventy-four percent responded. The test-retest correlations were all above 0.8, but while the variance of the absolute and arithmetical differences were acceptable and similar for QoL5 and QoL9, those for QoL1 were high.

Sensitivity

The ability of QoL1, QoL5, and QoL9 to detect differences in QoL are shown in Table V. It would need about 100 respondents in each group to detect a 3% difference in QoL between two groups, and about 10 to detect a 10% difference using QoL5 and QoL9, but substantially more respondents would be needed using QoL1 because of the relatively larger SD. Here, as has become scientific practice, the risk of a type I or II error were chosen to be 0.05 (two-sided) and 0.20; that is, one combines an 80% guarantee of correct detection with a 95% protection against spurious detection.

DISCUSSION

Global and generic quality of life measurements

The aim of introducing QoL measurement in clinical databases is to secure the quality of outcome concern-

Table III. *Internal Consistency. Correlations between individual items and total and partial QoL scores*

Questionnaire	Correlation between item and overall QoL score	The three partial QoL scores correlated with the overall QoL score
QoL5	Item 1 $r = 0.72$	Subjective QoL $r = 0.61$
	Item 5 $r = 0.68$	Existential QoL $r = 0.72$
	Item 6 $r = 0.63$	
	Item 7 $r = 0.58$	Objective QoL $r = 0.85$
	Item 8 $r = 0.70$	
QoL9	Item 1 $r = 0.60$	Subjective QoL $r = 0.81$
	Item 2 $r = 0.69$	
	Item 3 $r = 0.70$	
	Item 4 $r = 0.87$	Existential QoL $r = 0.74$
	Item 5 $r = 0.67$	
	Item 6 $r = 0.63$	
	Item 7 $r = 0.47$	Objective QoL $r = 0.78$
	Item 8 $r = 0.73$	
	Item 9 $r = 0.58$	

Table IV. *Reproducibility (reliability) of the three short quality of life measurements QoL1, QoL5, and QoL9. Test-retest at an interval of one month.* n = *37. Spearmann's correlation coefficient, absolute and arithmetical differences (5)*

	Correlation (95% CI)	Absolute difference		Arithmetical difference	
		Mean	(SD)	Mean	(SD)
QoL1	0.97 (0.92–0.99)	0.052	(0.23)	0.052	(0.23)
QoL5	0.82 (0.58–0.93)	−0.044	(0.17)	0.096	(0.14)
QoL9	0.88 (0.70–0.95)	−0.046	(0.14)	0.085	(0.13)

ing QoL. Age, sex, and disease(s) usually differ among the patients. The optimal QoL measurement must not depend on specific diseases, but should be global and generic to compare QL between patients, populations of patients, and the background population. However, this does not leave out the importance of including additional items of importance in the quality control of treatment.

The questionnaires use a theory of QoL that integrates most existing QoL theories. This theory secures a global and generic concept.

The ideal weighting of the three different QoL subscores is unknown. Instead of weighting the subscores according to which weighting best fits the measurements used for evaluation, we decided to weight the scores equally before analysis to secure the transparency of the method. Unfortunately, this weighting is also arbitrary but based on psychological and philosophical instead of empirical grounds.

Is it realistic to search for a theory-based QoL measurement consisting of a few items?

If it is, it would solve some major methodological problems about which items and domains to include, and how the individual items should be weighted. An illustrative example: Imagine a questionnaire for the evaluation of doctors' professional skills. How is the technique of suturing to be weighted? Which other questions must be asked? No matter how many items one includes, it would still leave uncovered areas. However, the introspective, analytical, and intellectual skills of human beings are superior to the capabilities of static questionnaires. These abilities could be used by asking more generalised and deep questions, for example: how do you consider your skills as a doctor? This would create an individually-weighted and deep, subjective evaluation. However, the sensitivity of one item is limited, and you may wish to make sure that central domains are considered. Consequently, one could ask: how does one consider one's communication, clinical judgement, or surgical skills—still using the introspective, analytical, and intellectual skills of human beings. However, without a theory and definition of doctoral skills, the selection and initial weighting of items become intuitive and incomplete. With a theory and definition, items included and weighting become logical. However, the logic of the weighting disappears by including many items.

Consequently, the development of theory-based short-item questionnaires solves some severe methodological problems in QoL research. However, if the few questions are not well targeted, the risk of short-item-based questionnaires is loss of external consistency (14).

Construct and criteria validity

The construct validity of the short questionnaires depends on acceptance of this integrated theory and definition (Fig. 1). By combining the theory with the items, and accepting that the theory describes QoL, the construct validity and weighting of the items seem valid.

Because the questionnaires were based on this theory, and not on the existing tradition of health-

Table V. *Sensitivity of the three new QoL measurements based on SD. The smallest number of respondents* (n) *in each group needed to detect a difference of 3%, 10%, or 20% in total QoL-score between two groups. The conventional two-sided probability of 0.05 was used in conjuction with an aimed power of 0.80*

Percentage difference in QoL score	QoL1 (*n*)	QoL5 (*n*)	QoL9 (*n*)
3% difference in QoL score	197	116	99
10% difference in QoL score	19	12	11
20% difference in QoL score	6	4	4
SD	15.1%	11.6%	10.7%

Dear Mr/Mrs/Miss

To evaluate the benefits of appointments and treatments in the health services, we would like you to answer a few questions about your quality of life.

Please consider the questions carefully before answering. Then draw a circle around the most suitable answer.

Q1. How do you consider your **physical health** at the moment ?

1 Very good
2 Good
3 Neither good or bad
4 Bad
5 Very bad

Q2. How do you consider your **mental health** at the moment?

1 Very good
2 Good
3 Neither good or bad
4 Bad
5 Very bad

Q3. How is your relationship with your **partner** at the moment ?

1 Very good
2 Good
3 Neither good or bad
4 Bad
5 Very bad /I do not have one

Q4. How are your relationships with your **friends** at the moment ?

1 Very good
2 Good
3 Neither good or bad
4 Bad
5 Very bad

Q5. How do you **feel about yourself** at the moment ?

1 Very good
2 Good
3 Neither good or bad
4 Bad
5 Very bad

Please make certain that you have answered **all** the questions. Thank you for your help.

Fig. 2. The QL5 and how to calculate the overall QoL. Objective QoL (from QI and QII on health) = (I + II)/2. Existential QoL (from QIII and QIV on relationships) = (III + IV)/2. Subjective QoL (from QV on wellbeing) = V. Overall QoL = (Objective QoL + Existential QoL + Subjective QoL)/3 = ((I + II)/2) + ((III + IV)/2) + V)/3. Roman numerals refer to the scored, encircled answer of the question with the same number, i.e.: if the respondent answers "2" in Q. 1, then I = (110 – (answer × 20))% = (110 − 2 × 20)% = 70%. Example: If the five answers obtained from Q.s 1–5 are 1, 2, 3, 4 and 5, respectively, the Overall QL = ((90% + 70%)/2 + (50% + 30%)/2 + 10%)/3 = 43.3%

related QoL, a low correlation with SIP (3, 15) and NHP (12, 13) was expected. However, concerning QoL5 and QoL9, the correlations were all above 0.5 concerning people who were unwell, but correlations were better with self-estimated QoL ($r = 0.68$–0.80). Self-estimated QoL is a validated, global, and generic questionnaire based on philosophical theories of life coming together in the integrative theory of QoL (18, 19, 20). Consequently, QoL5 and QoL9 correlate reasonably well with health-related QoL, and well with a global and generic QoL-measurement.

Reliability; internal consistency and reproducibility

Cronbach's α was calculated as described and recommended by McDowell and Newell (14). A large Cronbach's α indicates redundant interrogation, while a low value indicates a mixing of life elements that are uncorrelated, and so reflects things that should not be added unless they add up in some fundamental psychological sense. Consequently, a relatively poor Cronbach's α was expected. However, it was similar to the reference standards used, especially concerning people who were unwell, probably because the items in the formula favour many areas.

Another way of evaluating the internal consistency is Siegel's test, which correlates all the items with the overall QoL score (17). All correlations were above 0.5 apart from one. Consequently, QoL5 and QoL9 are homogeneous and consequently consistent.

Finally, the test-retest correlations were all above 0.8, which must be considered acceptable reproducibility.

From our own and others' experience we were expecting a relatively low response rate. Reproducibility studies require breaking anonymity, which we feared would have lowered the response rate even more. Consequently, we decided on a more reliable patient population. By selecting atherosclerotic patients, we could also achieve a positive side effect: to see whether the questions were also reasonable for an older atherosclerotic population. The high frequency of twice-completed questionnaires suggests that the questions are also reasonable and understandable for such people.

Sensitivity

Only about 100 respondents in each group would be needed to detect a 3% difference in QoL score between

two groups with the use of QoL5 or QoL9. However, whether such a small difference ever becomes of clinical interest is debatable. Clinically large and important differences—10% or more in overall QoL (see Fig. 2)—seem detectable with the use of only 10–20 respondents, which must be considered acceptable. The sensitivity would increase further through the use of paired data, which are often available in clinical databases.

CONCLUSIONS

Based on a QoL theory and the introspective, analytical, and intellectual skills of human beings, short global and generic QoL measurements were defined and developed (QoL1, QoL5, and QoL9).

Both QoL5 and QoL9 seem to have acceptable construct validity, external reliability, sensitivity, and internal reliability. The relatively low correlation with SIP is probably because SIP was designed to measure disease (3, 15), and seems to lose relevance among healthy people.

Whether QoL9 or QoL5 should to be recommended for application in clinical databases is difficult to answer. However, the various measurements of validity seem to suggest only marginal differences, except in the internal consistency of the subjective partial QoL-score. The shortest (QoL5) must therefore be recommended. Consequently, the items of QoL5 and the calculation formula have been released for public use without copyright in Fig. 2. This is how far we were able to go because of the many difficulties and obstacles connected with the resources, design, non-response, and need for anonymity. Further research is needed to evaluate the translated version and to test populations of patients and determine how they find the instrument.

ACKNOWLEDGEMENTS

We thank the 1991 Pharmacy Foundation of 1991, The Goodwill Foundation, The Foundation of Asta and Rosa Jensen, The Danish Heart Foundation, J. L Foundation, Foundation of Manager E. Danielsen and his wife, Emmerik Mayers Fellowship, Foundation of Frimodt-Heineken, The Foundation of Hede Nielsen Family, The Foundation of Petrus Andersens Fond, Grosserer C. F. Frederiksens Scholarship, Foundation of Else & Mogens Wedell-Wedellsborg, and IMK Almene Fond for financial support. We thank Dr. S. Vammen for collecting data for the reproducibility study.

REFERENCES

1. Aggernæs A. Livskvalitet (The quality of life). Copenhagen: FADL's Forlag, 1989.
2. Andrews FM. Research on the QL. Michigan: The University of Michigan, 1986.
3. Bergner M, Bobbit RA, Kressel S. The Sickness Impact Profile: Conceptual foundation and methodology for the development of a health status measure. Int J Health Serv 1976; 6: 393–415.
4. Bergner M. Quality of life, health status and clinical research. Med Care 1989; 27, Suppl 3: 148–156.
5. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; i: 307.
6. Breslin S. Quality of life: how is it measured and defined? Urol Int 1991; 46: 246–251.
7. Calman KC. The quality of life of cancer patients. In: Aaronson NK, Beckmann J, eds. Definitions and dimensions of quality of life. New York: Raven Press, 1987; 1–10.
8. Editorial. Quality of life. Lancet 1991; 338: 350–351.
9. Fitzpatrick R. Measuring the quality of life: from theory to practice. BMJ 1996; 313: 1341.
10. Henriksen BL. Livskvalitet (Quality of life). Copenhagen: GAD, 1992.
11. Holm P, Holst J, Olsen SB, Perlt B. Liv og kvalitet i omsorg og pædagogik (Life and quality in care and pedagogy). Herning: Systime, 1994.
12. Hunt SM, McKenna SP, Beckett EM, William J. Quantitative approaches to perceived health status: a validation study. J Epidemiology Community Health 1980; 34: 281–286.
13. Hunt SM, McKenna SP, William J. Reliability of a population survey tool for measuring perceived health problems: a study of patients with osteoarthrosis. J Epidemiology Community Health 1981; 35: 185–188.
14. McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. New York: Oxford University Press, 1987.
15. Pollard WE, Bobbit RA, Bergner M. The Sickness Impact Factor: reliability of a health status measure. Med Care 1976; 14: 146–155.
16. Sandøe P. QL and ethical priority. Copenhagen: Nyt Nordisk Forlag, 1992.
17. Siegel S. Nonparametric statistics for the behavioral sciences. Toyko: McGraw-Hill, 1956.
18. Ventegodt S. Resultater fra en befolkningsundersøgelse (Quality of Life in Denmark: Results from a population survey). Copenhagen: Forskningscentrets Forlag, 1995.
19. Ventegodt S. Measuring the Quality of Life: From Theory to Practice. Copenhagen: Forskningscentrets Forlag, 1996.
20. Ventegodt S. Livskvalitet hos 4500 31–33 årige (The Quality of Life of 4500 31–33-year-olds). Copenhagen: Forskningscentrets Forlag, 1996.
21. Walsh DL, Emrich LJ. Measuring cancer patients' quality of life: a look at physician attitudes. N Y State J Med 1988; 88: 354–357.
22 Wortis J. Quality of life (Editorial). Biol Psychiatry 1988; 23: 541–542.

Submitted April 06, 2001; submitted after revision August 07, 2001; accepted October 01, 2001

Address for correspondence:
J. S. Lindholt, M.D., Ph.D.
Frederik Vs Vej 15
3rd Floor, Room 21
DK-2100 Copenhagen
Denmark
Email: jslindholt@mail.tele.dk