

Comparing properties of the 15D and the EQ-5D in measuring health-related quality of life

OBJECTIVE To describe briefly the health state descriptive system and the valuation system of the 15D, and to compare its properties with the EQ-5D in terms of certain criteria such as feasibility, reliability, validity and sensitivity. **METHOD** One data set came from a Finnish population sample (n=359). Correlation co-efficients, multitrait-multimethod matrix and regression techniques were used for analysis. Another data set consisted of Norwegian patients with chronic obstructive pulmonary disease (n=59). In this sample the scores generated by the 15D, EQ-5D based on British Time Trade-Off (TTO) valuations, TTO on own present health state and Standard Gamble (SG) on own present health state were compared. Spearman's rank correlations, effect size and responsiveness statistic were used for analysis. **RESULTS** The 15D was equally acceptable to the EQ-5D, but superior to it in reliability, discriminatory power and responsiveness to change. **CONCLUSIONS** The 15D is preferable in studies where the properties of reliability, discrimination and responsiveness are important.

Although the measurement of health-related quality of life (HRQoL) in health care has become popular only in recent years, the development of many instruments for this purpose has a longer history. Concerning European generic (non-disease specific) and preference-based, single index number instruments, the development of what is now known as the 15D started in the late 1970s. Since 1987 the EuroQol Group has been working on the development and testing of a simple, self-administered, standardized, non-disease specific instrument now known as EQ-5D.

The present EQ-5D health state descriptive system consists of five dimensions: mobility, self-care, usual activities, pain or discomfort, anxiety or depression. Each dimension is divided into three levels, so the instrument defines 243 health states. The main purpose of the EuroQol Group was to create with this descriptive system a set of standard health states and with them to explore the question of whether health state valuations are similar across a number of European countries. A visual

analogue scale (rating scale) was selected as the standard method for eliciting valuations from the general public by using a self-administered questionnaire in a postal survey.^{1,2}

Another idea of the EuroQol Group was that due to the simplicity and easy applicability, the instrument could always be included in health program evaluation studies in addition to whatever other HRQoL measure, disease-specific or generic, the researchers might be interested in applying. Thus the EQ-5D would not be itself an adequate stand-alone measure, but it would complement other forms of HRQoL measure and serve as a linkage tool between different measures and circumstances. In such a capacity the EQ-5D would guarantee some degree of comparability across different studies using different HRQoL measures.¹

The purpose of this paper is to describe briefly the health state descriptive system and the valuation system of the 15D, and to compare its properties with the EQ-5D in terms of several criteria such as feasibility, relia-

H. Sintonen

Department of Public Health,
University of Helsinki, Finland

Σύγκριση των κλιμάκων 15D και EQ-5D
στη μέτρηση της σχετιζόμενης
με την υγεία ποιότητας ζωής

Περίληψη στο τέλος του άρθρου

Key words

EQ-5D, 15D
Quality of life
Sensitivity
Validity

bility, validity and sensitivity. The comparison is based partly on results from a population sample, partly from a patient group. It was anticipated that the study may also indicate how the EQ-5D might perform as a linkage tool.

When developing the 15D the basic purpose was to create a generic, multi-dimensional, standardized, self-administered measure of HRQoL, which could be used primarily as a single index score measure, but also as a profile measure. The idea was to combine the advantages of a profile and a preference-based measure. The first 12-dimensional version was published 1981³ and it was revised into the first 15-dimensional version in 1986.⁴ In this study the latest version from 1992 was used. It consists of the following 15 dimensions: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort and symptoms, depression, distress, vitality and sexual activity. Each dimension is divided into five levels so the health state descriptive system defines an enormous number of health states. The 15D thus covers the physical, psychological and social aspects of health as defined by the WHO. The 15D is more comprehensive in sampling items for the construct of HRQoL than any other instrument of its type and therefore it lends itself to more accurate inferences and possesses a higher content validity.⁵

To generate a single index score on a 0–1 scale (0=being dead, 1=full health) for all possible states, an additive 3-stage valuation model, based on the multi-attribute utility theory, is applied.

For this model, valuations are elicited from the general public by using ratio scales with self-administered questionnaires in a postal survey. The preference weights thus derived reflect the relative importance of different dimensions and the distance between various levels within each dimension from the viewpoint of HRQoL.^{3,6}

MATERIAL AND METHOD

The properties of the 15D and EQ-5D health state descriptive systems are compared on the basis of results from a Finnish population sample. For the 15D valuation study five random samples (n=500 each) of the Finnish population ages >16 years were drawn from the National Population Register. In the stratified sampling the elderly (aged 65+ years) were over-represented to compensate for their lower absolute number in population and a possibly higher non-response rate. One of the samples received a questionnaire which included the EQ-5D and 15D descriptive systems in addition to the background data (age, gender, education, whether experienced serious illness self, in family and when treating others, whether a person has at present an illness or impairment and its duration).

This sample is considered in this paper. The 15D single index score was assigned to the respondents of this sample by using the preference weights derived in this valuation study.⁶ For analyses the sample was made compatible with the age and gender structure of the whole adult population⁷ by appropriate weighting.

The EQ-5D single index score on a 0–1 scale for the respondents was assigned by using the Finnish “tariff” generated with regression analysis from the direct valuations of 42 states in another Finnish postal survey. For the survey a random sample (n=4,000) of the Finnish population aged >16 years was drawn from the National Population Register.⁸

The feasibility of the health state descriptive systems is compared by looking at the completion rates by dimensions.

Construct validation involves gathering external empirical evidence, convergent or discriminant, so that meaningful inferences can be made with the measure. To show convergent validity the measure should correlate highly with other variables and other measures of the same construct, to which it should correlate on theoretical grounds. Discriminant validity implies that the measure should not correlate with dissimilar, unrelated variables or measures.⁹

Multitrait-multimethod matrix¹⁰ is used to look at convergent and discriminant validity simultaneously. It is expected that the Pearson correlations of 15D with comparable 15D and EQ-5D dimensions are higher (convergent validity) than with non-comparable ones (discriminant validity).

To exhibit convergent evidence for the single index scores, the different sets of scores for the respondents’ own health state were correlated (Pearson). In addition, regression techniques were used to explore the relationship between the value sets more closely.

As another test for convergent evidence for the single index scores, extreme groups comparison with t-test is used to test the following hypotheses: (a) There is a significant difference in the EQ-5D and 15D scores between the elderly (65+ years), middle-aged (36–64 years) and young people (17–35 years) with the scores being lower, the older people are. (b) People reporting an illness or impairment have a significantly lower mean score on EQ-5D and 15D than people without an illness or impairment.

Sensitivity of a measure entails two aspects. First, the ability to distinguish between individuals and groups in different health states cross-sectionally (discriminatory power) and second, to detect changes in individuals or groups over time (responsiveness to change in health status).¹¹

The discriminatory power of the health state descriptive systems is compared in the population sample. Patrick and Erickson¹² mention three criteria for evaluation. First, the ability to detect health problems, especially in a relatively healthy population. Second, the ability to detect improving health among quite healthy people, and to avoid the “ceiling” effect of having no better health state to go to. Third, the ability to detect worsening health among people who are already quite ill, that is, to avoid the “floor” effect of having no worse health state to go to.

The discriminatory power is examined empirically by looking at the percentages of respondents who score the “ceiling” for different dimensions and the measure as a whole. The corresponding percentages at the “floor” indicate the range of health states used.

The properties of the HRQoL scores generated by the valuations systems are also compared in the light of the results obtained by Stavem¹³ in patients with chronic obstructive pulmonary disease. He compared the scores generated by the 15D, the EQ-5D based on British Time Trade-Off (TTO) valuations¹⁴ (EQ-TTO), TTO on own present health state and Standard Gamble (SG) on own present health state. In addition he correlated these scores with a number of clinical measures in the patients.

RESULTS

Properties of the health state descriptive systems

Response and completion rates. In the population sample under study, the response rate was 72% giving a final sample of 359 observations. The completion rates by dimensions were 96–99% for the 15D (except sexual activity which was 90%) and 97–98% for EQ-5D. Regarding completion rates, the acceptability of 15D is thus comparable to EQ-5D. The lower completion rate for the dimension of sexual activity may indicate that this dimension is slightly less acceptable than the others.

Construct validity. The multitrait-multimethod matrix in table 1 shows that the correlations of 15D with comparable EQ-5D dimensions (in italics) are consistently higher than the correlations with non-comparable scales measuring dissimilar attributes. This is a pattern that scales with convergent and discriminant validity are expected to exhibit thus providing solid evidence for the construct validity of these measures.

Discriminatory power. Table 2 shows the discriminatory power of the 15D and EQ-5D by comparable dimensions. Apart from mobility, the much higher percentages at the “ceiling” for EQ-5D suggest that it has less discriminatory power than comparable 15D dimensions. Moreover, it must be borne in mind that the remaining 9–10 dimensions of 15D provide an extra sensitivity reserve at both the ceiling and the floor. Its significance can be seen by looking at the discriminatory power of the measures as a whole. The EQ-5D classified 50.3% of respondents as “healthy”, the 15D only 19.7% (none were at the “floor” on either method). Thus the discriminatory power of the 15D is much better than that of EQ-5D in the general public.

Properties of the health state valuation systems

The average 15D score in the population sample was 0.914 (SD=0.096, min=0.45) and average EQ-5D score was 0.822 (SD=0.195, min=0.12). The means represent the average HRQoL scores of the adult Finnish population.

The correlation of the 15D scores with the EQ-5D scores was quite high, $r=0.757$ ($P<0.000$). This provides convergent evidence of construct validity for the scores. Below are shown the best regression equations for converting the EQ-5D scores into 15D scores and vice versa. The fit between the EQ-5D and 15D scores is reasonably good.

$$15D = 0.222 + 1.443 * EQ - 0.694 * (EQ)^2$$

$$\text{adjusted } R^2 = 0.635 \quad (P < 0.000)$$

$$EQ = 0.789 + 1.999 * 15D - 1.810 * (15D)^2$$

$$\text{adjusted } R^2 = 0.647 \quad (P < 0.000)$$

Table 3 shows the average EQ-5D and 15D scores in different age groups and for respondents with and without a self-reported illness or impairment. The average scores

Table 1. Multitrait-multimethod matrix of correlations for comparable EQ-5D and 15D dimensions.

	EQMOB	EQUACT	EQPAIN	EQMOOD	MOB	UACT	D & S	DEPR	DISTR
EQMOB	1.00								
EQUACT	0.59	1.00							
EQPAIN	0.57	0.49	1.00						
EQMOOD	0.17	0.29	0.23	1.00					
MOB	0.77	0.63	0.52	0.22	1.00				
UACT	0.65	0.76	0.54	0.24	0.71	1.00			
D & S	0.46	0.42	0.65	0.31	0.46	0.45	1.00		
DEPR	0.15	0.25	0.20	0.67	0.17	0.22	0.36	1.00	
DISTR	0.21	0.26	0.26	0.57	0.20	0.29	0.43	0.63	1.00

EQ-5D dimensions: EQMOB=mobility, EQUACT=usual activities, EQPAIN=pain or discomfort, EQMOOD=anxiety or depression. 15D dimensions: MOB=mobility, UACT=usual activities, D & S=discomfort and symptoms, DEPR=depression, DISTR=distress

Table 2. The discriminatory power of comparable 15D and EQ-5D dimensions in terms of “ceiling” and “floor” effects.

Dimensions	At ceiling (%)	At floor (%)
Mobility/15D	82.0	–
Mobility/EQ-5D	78.9	0.2
Usual activities/15D	73.3	1.0
Usual activities/EQ-5D	79.0	1.7
Depression/15D	60.0	1.2
Distress/15D	62.4	0.8
Anxiety/depression/EQ-5D	85.1	1.1
Discomfort/symptoms/15D	40.5	0.4
Pain or discomfort/EQ-5D	58.1	1.8

are significantly different in the age groups on both measures, and as expected the older the respondents get the lower score. The HRQoL scores are also significantly different in the groups with and without an illness or impairment. These comparisons provide further evidence of cross-sectional convergent construct validity.

Reliability. In patients with chronic obstructive pulmonary disease the test-retest reliability of the 15D over 14 days was clearly better than that of EQ-TTO (this reliability was not tested for TTO and SG). The 15D scores deviated clearly less from (=agreed better with) the TTO and SG scores than the EQ-TTO scores and had a higher rank correlation with the TTO scores than the EQ-TTO. The 15D showed moderate to high rank correlations with commonly used clinical measures of symptoms, lung function and exercise capacity and for most measures, higher than EQ-TTO, TTO and SG.¹³

Sensitivity. The 15D was better at discriminating between groups after reported global change in health status than EQ-TTO, indicating thus better responsiveness to change. Moreover, the 15D had higher values for effect size and responsiveness statistics for the two groups of patients reporting a global change in health status. On the other hand the 15D scores turned out to be more stable than the EQ-TTO scores over a 12 month period in patients, who reported no change in global health status, indicating thus a better reproducibility for the 15D.¹³

Validity. There is no gold standard for valuation system, that is, for how to measure the values and from whom. To be valid for QALY calculations, the values should reflect a reasonable trade-off between quality and length of life. If SG or TTO valuations on own health are taken as the gold standard, the results of Stavem¹³ showed that the 15D scores agreed better with them than the EQ-TTO scores, and are thus more valid for QALY calculations.

DISCUSSION

In the light of the completion rates, the EQ-5D and 15D are equally acceptable. The multitrait-multimethod matrix and comparison of average scores in different age groups and in groups with and without a self-reported illness or impairment provide clear convergent and discriminant evidence of construct validity. The convergent validity correlations of the 15D were quite high being in the range of 0.57–0.77. In their review of health measures McDowell and Newell¹⁵ found that the validity coefficients fell typically between 0.20 and 0.60.

The empirical evidence suggests that in discriminatory power, on roughly comparable dimensions and as a whole measure, the 15D is superior to the EQ-5D. The same applies to responsiveness to change. Also Stavem¹³ concluded that “the 15D instrument has many attractive properties when compared to the EQ-TTO method, including a better reliability and responsiveness”. This is as expected, since EQ-5D was not originally designed to be a sensitive stand-alone measure, but a simple linkage tool between more comprehensive measures such as, for example, the 15D.

In relation to the 15D, the EQ-5D can be quite useful and satisfactory as a linkage tool for two reasons. First, it takes clearly less time to complete the EQ-5D health state descriptive system than that required to complete the 15D (15D about 5–10 minutes) so these health state descriptive systems can be included in the same questionnaire with very little extra burden to the respondents. Even if this is not done, the regression equations suggest that the EQ-5D scores can be converted to 15D scores or vice versa with a reasonable accuracy.

Table 3. The average EQ-5D and 15D scores in different age groups and for respondents with and without a self-reported illness or impairment (95% CI in parentheses).

	Age in years			Illness or impairment	
	17–35	36–64	65+	No	Yes
15D	0.961 (0.953–0.966)	0.914 (0.897–0.932)	0.812 (0.790–0.833)	0.954 (0.945–0.963)	0.836 (0.818–0.854)
EQ-5D	0.904 (0.844–0.932)	0.830 (0.749–0.867)	0.646 (0.613–0.679)	0.919 (0.897–0.941)	0.665 (0.638–0.691)

ΠΕΡΙΛΗΨΗ

Σύγκριση των κλιμάκων 15D και EQ-5D στη μέτρηση της σχετιζόμενης με την υγεία ποιότητας ζωής

H. SINTONEN

Department of Public Health, University of Helsinki, Finland

Αρχαία Ελληνικής Ιατρικής 2001, 18(2):156-160

ΣΚΟΠΟΣ Το αντικείμενο της έρευνας αυτής είναι η περιγραφή του συστήματος αξιολόγησης της κατάστασης υγείας 15D και η σύγκρισή του με το σύστημα EQ-5D με βάση διάφορα κριτήρια, όπως η σκοπιμότητα, η αξιοπιστία, η εγκυρότητα και η ευαισθησία. **ΥΛΙΚΟ-ΜΕΘΟΔΟΣ** Χρησιμοποιήθηκαν δύο δείγματα. Το πρώτο δείγμα προήλθε από γενικό πληθυσμό της Φινλανδίας (n=359) και για την ανάλυσή του χρησιμοποιήθηκαν συντελεστές συσχέτισης, πολυδιάστατη μήτρα και τεχνικές παλινδρόμησης. Το δεύτερο δείγμα προήλθε από ασθενείς της Νορβηγίας με χρόνια πνευμονοπάθεια (n=59) και σε αυτό το δείγμα εξετάστηκαν τα αποτελέσματα από το 15D και το EQ-5D βασισμένο στις ΤΤΟ αξιολογήσεις. **ΑΠΟΤΕΛΕΣΜΑΤΑ** Το 15D ήταν εξίσου αποδεκτό με το EQ-5D, αλλά ανώτερο σε αξιοπιστία, σε δυνατότητα διάκρισης και σε ανταποκρισιμότητα σε αλλαγές. **ΣΥΜΠΕΡΑΣΜΑΤΑ** Το 15D είναι προτιμητέο στις μελέτες όπου όλες οι παραπάνω παράμετροι είναι απαραίτητες.

Λέξεις ευρετηρίου: Αξιοπιστία, Ευαισθησία, Ποιότητα ζωής, 15D, EQ-5D

References

1. EUROQOL GROUP. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990, 16:199-208
2. BROOKS R, WITH THE EUROQOL GROUP. EuroQol: the current state of play. *Health Policy* 1996, 37:53-72
3. SINTONEN H. An approach to measuring and valuing health states. *Soc Sci Med* 1981, 15C:55-65
4. SINTONEN H, PEKURINEN M. A fifteen dimensional measure of health-related quality of life (15D) and its applications. In: Walker SR, Rosser RM (eds) *Quality of Life Assessment. Key issues in the 1990s*. Kluwer, Dordrecht, 1993:185-195, 467-470
5. SINTONEN H. The 15D-measure of health-related quality of life. I. Reliability, validity and sensitivity of its health state descriptive system. Working Paper 41. National Centre for Health Program Evaluation, Melbourne, 1994
6. SINTONEN H. The 15D-measure of health-related quality of life. II. Feasibility, reliability and validity of its valuation system. Working Paper 42. National Centre for Health Program Evaluation, Melbourne, 1995
7. STATISTICS FINLAND. Statistical yearbook of Finland 1993. Printing Centre, Helsinki, 1993
8. OHINMAA A, SINTONEN H. Inconsistencies and modelling of the Finnish EuroQol (EQ-5D) preference values. In: Greiner W, Graf u.d. Schulenburg J-M, Piercy J (eds) Plenary meeting. Hannover 1998. Discussion Papers. Uni-Verlag Witte, Hannover, 1999:57-74
9. STREINER DL, NORMAN GR. *Health Measurement Scales: A practical guide to their development and use*. Oxford University Press, Oxford, New York, Tokyo, 1989
10. CAMPBELL DT, FISKE DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959, 56:81-105
11. KIRSHNER B, GUYATT GH. A methodological framework for assessing health indices. *J Chron Dis* 1985, 38:27-36
12. PATRICK DL, ERICKSON P. Assessing health-related quality of life for clinical decision making. In: Walker SR, Roser RM (eds) *Quality of Life Assessment. Key issues in the 1990s*. Kluwer, Dordrecht, 1993:11-63
13. STAVEM K. Reliability, validity and responsiveness of two multi-attribute utility measures in patients with chronic obstructive pulmonary disease. *Qual Life Res* 1999, 8:45-54
14. WILLIAMS A. The measurement and valuation of health: A chronicle. Discussion Paper 130. Centre for Health Economics, York, 1995
15. McDOWELL I, NEWELL C. *Measuring Health: A guide to rating scales and questionnaires*. Oxford University Press, New York, Oxford, 1987

Corresponding author:

H. Sintonen, Department of Public Health, P.O. Box 41, 00014 University of Helsinki, Finland