Arch Public Health
2003, 61, 201-214

Measuring health state preferences in Belgium with the EQ-5D: a pilot survey in health care workers

by

Cleemput I.[1], De Geest S.[2], Vanrenterghem Y.[3], Kesteloot K.[1, 4]

Abstract

The EuroQol valuation instrument is an internationally standardised instrument for the measurement of health state preferences in the general public. The EuroQol valuation instrument, or the EQ-5D, has never been tested in a Belgian population. The aim of this pilot study was to explore the feasibility, consistency and reliability of using the EQ-5D valuation instrument in the Belgian population on the basis of a survey in a Belgian sub-population.

The questionnaire was distributed to 274 health care workers. Feasibility was assessed by response rate, perception of difficulty, and time needed to fill out the questionnaire. For consistency, examined was whether pairs of health states with an inherent logical order were valued

Address for correspondence: Dr. Irina Cleemput – Terlinden 56 – B-1785 Merchtem – Tel. +32 52 57 01 09 – Fax. +32 52 57 01 10 – E-mail: Irina.Cleemput@pandora.be

[1] K.U.Leuven, Centre for Health Services and Nursing Research (Current affiliation: HEDM, Brusselsesteenweg 91, 1860 Meise)

[2] K.U.Leuven, Centre for Health Services and Nursing Research (Current affiliation: Institute of Nursing Science, University of Basel)

[3] Department of Nephrology, University Hospitals Leuven

[4] C.F.O., University Hospitals Leuven

accordingly. Valuations of states appearing twice in the questionnaire were compared to evaluate reliability. Given the non-normality of the data, non-parametric tests were used for statistical analyses.

A response rate of 49% was obtained. Average consistency was 97.33%. Reliability ranged from 70 to 80%. Given the higher response rate and the higher level of consistency in our sample of health professionals as compared to results from population surveys in other countries, we conclude that it is worthwhile to use the EQ-5D valuation questionnaire in Belgium to derive social health state preference values from the general public.

Keywords: *quality of life, health status index, visual analogue scale, validity and reliability, feasibility*

Introduction

The increasing attention in policy decisions to health-related quality of life as the outcome of interest for health interventions, requires the use of a generic outcome measure. There is a general agreement that the quality of life perceptions of the general public should be given important weight in the evaluation of outcome, as it is precisely the general population that is paying *and* benefiting from care (1). Many countries have tried to develop a reference set of valuations for general health state descriptions that can be used in outcome assessment of health interventions. In Belgium, no such set is available.

A reference set of health state valuations is only useful if the instrument used for this purpose is feasible, valid, reliable and generates consistent valuations. Feasibility refers to the practicality and acceptability of an instrument (2). Validity means that the instrument actually measures the concept it claims to measure (3). In the context of quality of life measurement, validity refers to the extent to which the method fulfills its predictions (i.e. construct validity) and the extent to which the instrument captures the whole range of relevant aspects of the phenomenon under study (i.e. content validity) (4). Reliability refers to the stability of valuations, either over time or within respondents (3). Finally, consistency refers to the extent to which the method entails logical orderings of health states (3).

This paper reports on a survey performed in a group of health professionals to explore the feasibility, consistency, reliability and validity of the EQ-5D valuation questionnaire in deriving health state preferences from the Belgian general population. The study is a pilot project for future

population research in Belgium. The hypothesis is that a survey in health professionals with the EQ-5D shows higher feasibility, consistency, reliability and validity as compared to literature data on general population surveys. If the hypothesis cannot be rejected, it is worthwhile to consider a larger population survey in Belgium.

Methods

The EQ-5D valuation questionnaire

The EQ-5D valuation questionnaire is a generic instrument for the measurement of health state preferences. The instrument was developed by a multidisciplinary group of researchers, the EuroQol Group (5), from fields of economics, mathematics, medicine, nursing, philosophy, psychology and sociology. The aim of the EuroQol Group was to devise an instrument to collect valuations for a generic set of health state descriptions from the general public and use these valuations for reference purposes in cost-effectiveness analyses. The acronym 'EQ-5D' stands for EuroQol 5 dimensions, referring to the five-dimensional classification system for health states used in the instrument. Each dimension (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) has three levels of severity (no, some and severe problems), theoretically allowing for 243 (3^5) different health states. Health states are given a 5-digit code, in which the digits represent the level of severity on the respective dimensions.

The valuation questionnaire consists of a limited selection of the 243 health states that can be described with the EQ-5D because it has previously been shown that it is psychometrically not feasible to let people value more than 16 health states during one interview or postal survey (6). Fourteen different hypothetical health states had to be valued on a thermometer-like visual analogue scale (VAS)[1], ranging from "worst imaginable health state" to "best imaginable health state". The states are presented on two pages. Two health states, 11111 and 33333 appear on both pages, to act as common reference points (7). "Death" is valued last by drawing a horizontal line over the VAS on both pages. The valuation of "death" allows for the calibration of the "raw" VAS scores of all other states on a scale from 0 (=death) to 100 (=perfect health) (8). The first page of the questionnaire is presented in Figure 1.

[1] The VAS is only one technique to measure health state preferences. In principle, the EQ-5D valuation questionnaire can be used in combination with any preference measurement technique (e.g. TTO or SG) and is not tied to VAS.

Fig. 1: Page 1 of the EQ-5D valuation questionnaire

The test-retest reliability of the EQ-5D with VAS as a measurement instrument for health state preferences is acceptable, as shown by studies in different patient populations and the general public (1;9-17), with mean internal consistency reliability coefficients (intra-class correlation coefficients, ICC) ranging from 0.65 to 0.93 for the health state valuations.

Content validity of the 5-dimensional descriptive system is based on the work and experience of researchers from different disciplines. The developers of the EuroQol instrument relied on existing instruments for the measurement of health-related quality of life, such as the Quality of Well-Being scale, the Nottingham Health Profile, the Sickness Impact Profile and the Rosser Index (5). Given its purpose to be useful in economic evaluations, the aim was to develop a generic instrument that

could easily be translated into one single index value for each health state. Simplicity rather than comprehensiveness was the main objective (7). This purpose was accepted as being valuable in itself. Therefore, content validity was not further analysed in this study.

Concurrent validity of the EQ-5D valuation instrument with VAS can only be described in terms of performance in comparison with a 'golden standard' for health state preference measurement. The problem is that no 'golden standard' for health state preference measurement is available against which the VAS could be tested. This is a general problem for all health state preferences research. None of the two commonly used alternatives, the Time Trade-Off (TTO) and the Standard Gamble (SG), can be regarded as being 'golden standards' (18) and therefore, this analysis is not performed.

Sample

Dutch-speaking health care workers on dialysis wards from 5 Belgian hospitals[2] were surveyed on their health state preferences with the EQ-5D valuation questionnaire with visual analogue scale. All health professionals working on the ward were contacted, without exception. Given that all five hospitals were located in the Flemish part of Belgium, none of the contacted health care workers was exclusively French speaking, therefore not necessitating the use of a French version of the EQ-5D, which was non-existing at the time of the study. Two-hundred and seventy-four nurses and physicians were asked to fill out the questionnaire and send it back to the investigator. The survey was anonymous.

Statistical analysis

A Kolmogorov-Smirnov test revealed that both "raw" and calibrated valuation data and consistency rates were highly skewed. Therefore, non-parametric tests were used to analyse the data. Differences in valuations between subgroups were tested by means of the Mann-Withney U test for comparisons between two groups (e.g. men and women) or by means of the Kruskal-Wallis test for comparisons between multiple groups (e.g. different age groups). Significance level was set at 5%.

[2] O.L.V. Ziekenhuis Aalst, Heilig Hartziekenhuis Roeselare, Sint Jansziekenhuis Genk, Virga Jesseziekenhuis Hasselt, Universitaire Ziekenhuizen Leuven.

Results

Sample characteristics

Table 1 lists the characteristics of the 121 respondents that yielded a usable response. As can be expected from a sample drawn mainly from a population of nurses, the majority of the respondents is female (78%). Mean age of the respondents was 39 years. As most subjects were registered nurses (N=109), the majority of the respondents had a higher non-university degree.

TABLE 1
Sample characteristics

Variable	Value
Age (n=119), mean (s.d.) in years	38.6 (9.2)
Gender (n=121)	
Male	27 (22.3%)
Female	94 (77.7%)
Smoking status (n=119)	
Current smoker	25 (21.0%)
Ex-smoker	16 (13.4%)
Non-smoker	78 (65.5%)
Mean (s.d.) number of years working as a health care worker (n=120)	16.9 (9.4)
Function (n=121)	
Nurse	109 (89.3%)
Physician	8 (6.6%)
Other	4 (3.3%)
Education (n=114)	
Higher education but not university	105 (92.1%)
University	9 (7.9%)

Feasibility

A total of 274 questionnaires was distributed to health professionals. Of the 134 questionnaires returned (48.9%), 6 (4.5%) were blank and 4 respondents did not perform the valuation task.
Traditional indicators for feasibility are the percentage of returned questionnaires filled out in a complete and valid manner (67.2%) and the number of valid and complete responses as a percentage of all questionnaires distributed (32.8%) (Table 2).

TABLE 2
Rates of return

	Number and/or rate
Distributed questionnaires (total sample), n (%)	274 (100%)
Returned questionnaires, n (%)	135 (49.2%)
of which: Blank or equal, n	*10*
1-3 missing VAS scores, n	*27*
4-6 missing values, n	*3*
7-11 missing values, n	4
Complete response (18 VAS scores), n	90
as % of total number of distributed questionnaires	32.8%
as % of returned questionnaires	67.2%

A third indicator of feasibility is task difficulty (Table 3). Seventy-four percent of the respondents rated the task as very difficult or fairly difficult. Respondents spent on average 24 minutes on the valuation task (self-report). The results are less optimistic than those found in literature (19), in which only 58% to 61% found the questions difficult and respondents used 20 minutes on average to complete the questionnaire (20).

TABLE 3
Task difficulty

Characteristic	**Number (%)**
Difficulty (n=117)	
Very difficult	12 (10.3%)
Fairly difficult	75 (64.1%)
Fairly simple	25 (21.4%)
Very simple	5 (4.3%)
Time needed to fill out questionnaire in minutes (n=117), Mean (s.d.)	24.2 (16.8)

Consistency

For the evaluation of consistency in health state valuations, examined is whether health states that are logically worse are valued lower than states that are logically better according to their description and vice versa. Inconsistent responses are, however, not necessarily a reflection of respondents' misinterpretation of the valuation task. Some researchers argue that inconsistencies might, for some respondents, have a fairly logical explanation. This is exactly the reason why in population studies, respondents with inconsistent responses are not a priori excluded from the data-set. Inclusion/exclusion of inconsistent

responses depends on whether these inconsistent valuations have a minor/major impact on the social health state valuations.

Seventy-five pairs with an inherent logical order were identified in the questionnaire. The consistency rate is defined as the percentage of all logical pairs that is valued as being consistent.

In contrast to Dolan and Kind (21), we did not find significant differences in consistency rates related to age differences. Although older persons are more prone to inconsistencies in valuations than younger persons, the difference between age groups is not significant according to the Chi-square test (p=0,27) (Table 4). However, it should be noted that the age distribution of our sample was very compact, which may explain the lack of significance. Neither sex nor smoking behaviour of respondents was related to the inconsistency level. This finding is in line with the results of other studies reported in the literature (21). Eighty percent of the respondents had a consistency rate of more than 90%. There were no inconsistencies in median VAS valuations on the group level.

TABLE 4
Task difficulty

Characteristic	**Medians (P_{25} – P_{75})**
Age	
20-29 (n=22)	98.7% (94.7-100)
30-39 (n=42)	98.7% (96.0-100)
40-49 (n=39)	97.3% (89.3-100)
50-60 (n=12)	94.0% (91.3-99.7)
Overall (n=115)*	97.3% (94.7-100)

* three respondents did not state their age

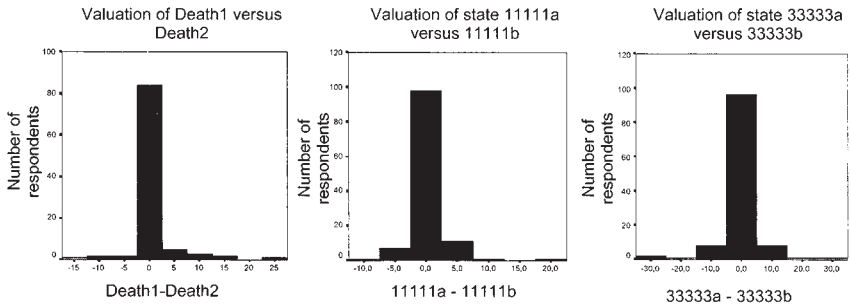
Reliability

As for the reliability test, it was examined to what extent health states that appear twice in the questionnaire are rated equally (split-test reliability). Test-retest reliability could not be tested as the survey was anonymous and responders could not be identified for re-test. According to the literature, the EuroQol valuation instrument has good test-retest reliability when it was used in the general population (1,13).

Figure II shows the frequency distributions of differences between valuations of identical health states. For states 11111, 33333 as well as death, the distributions strongly peak at 0, meaning that the majority of respondents valued the health states equally on page 1 and page 2 of

the questionnaire. Split-test reliability of the questionnaire lies between 70 and 80%. On the group level, median valuations for states 11111, 33333 and death on both pages were equal. State 11111 was assigned a value of 95, state 33333 and death a value of 10 on a scale from 0 (worst imaginable health state) to 100 (best imaginable health state).

Fig. 2: Frequency distribution of differences in valuations of identical health states



Validity

Construct validity of questionnaires is usually tested on the basis of expected relationships derived from the literature. Literature was reviewed in search for established relationships between respondent-related variables and health state valuations. Most studies find no significant impact of age and sex on health state valuations (1,12), although exceptions for a subset of states are possible (12,22). Social class, education and home ownership seem to have an impact on health state valuations in one study (1), education only in another (1,22). The impact of current experience of illness is ambiguous. Some authors report an impact of experience of illness on the valuations of a small number of health states(1,22), whereas others find an impact for a majority of states.(23,24). This summary of evidence from literature with respect to potential relationships between respondent characteristics and health state preferences illustrates the absence of valid hypotheses against which the construct validity of valuation instruments can be tested. Further validity tests were therefore considered inappropriate (1,23).

Discussion

This paper reports on a pilot project exploring the feasibility, consistency and reliability of using the EQ-5D valuation instrument in the Belgian population to develop a reference set of health state valuations.

Although our sample of respondents is not representative of the general population –our respondents were mostly highly educated and an overwhelming majority was female- our results were encouraging. The response rate, although disappointing, is higher than in population surveys with a postal EQ-5D valuation questionnaire in other countries (UK: 23.1% (5), The Netherlands: 37% (5), Sweden: 20.8% (20), Norway: 27% (19), Australia: 27% (25)). The higher response rate in our sample compared to population samples in other countries does not prove that the instrument is useful for the measurement of health state preference values in the entire Flemish population. The response rate is still rather low. It could be argued that this is a reflection of the mode of administration (i.e. postal), rather than of the feasibility of using the instrument in Belgium. However, postal survey is inexpensive compared to interviews, which typically give higher response rates (26). Yet, this is not a guarantee for success: interview-based surveys may show worse results in terms of consistency in the valuations. In an interview-based Catalan VAS survey, inconsistencies were found in 26% of the responses, which is a much higher inconsistency rate than the one found in our study. Evidence suggests that interviews may actually lead to more inconsistencies than postal surveys because the latter are likely to attract more respondents who are capable of giving consistent answers (26). Respondents that do not feel confident to give consistent answers might not return the questionnaire at all (21). The consistency rates found in our study concur with the result found by Dolan and Kind (21), who reported a median inconsistency rate of 2.7% for a self-completed postal EuroQol survey.

The split-test reliability of the results obtained from the valuation instrument was satisfactory, with levels between 70 and 80%. As the survey was anonymous, it was impossible to examine test re-test reliability, as well as non-response bias.

Due to homogeneity of the subjects in our sample, construct validity could not be tested. A similar self-completed postal survey in health professionals in New Zealand has found striking similarities, however, between health professionals' valuations and valuations from the British general public (27). This may be an indication of the fact that health professionals' valuations are not necessarily different from the valuations of the general public. Whether the valuations from our limited –clearly not representative- sample of health professionals deviate substantially from the valuations of the general public, needs more investigation.

The anonymity of the survey precluded a re-test analysis and an analysis of potential non-response bias. Especially the latter may be considered a major problem for the generalizability of the results to the population of health care workers in specific and to the general population in general. On the other hand, uncertainty about the potential difference between health state valuations of non-responders versus responders is a general problem for all anonymous postal surveys, whether in specific population groups or the population as a whole. Previous EuroQol studies have found that a postal EQ-5D valuation survey is not very sensitive to selection bias by non-response as far as the health state valuations are concerned (8). Differences in demographic characteristics of non-responders and responders only form a threat to generalizability if health state valuations are indeed influenced by these variables. As discussed earlier, evidence for age-, sex-, social class- and education-related differences in health state valuations is mixed. Whether or not a difference is found seems to depend on the country in which the study is performed.

Conclusion

The results of this pilot project revealed that it is worthwhile to perform a larger population survey with the EuroQol health state valuation instrument in Belgium. The results show better consistency and reliability than in population studies from other countries. Whether these results are maintained in a larger population study needs to be investigated.

Samenvatting

Het EuroQol waarderingsinstrument is een internationaal gestandaardiseerd instrument voor de meting van preferenties met betrekking tot gezondheidstoestanden in de algemene populatie. Het EuroQol waarderingsinstrument, of de EQ-5D, is tot op heden nooit gebruikt in een Belgische populatie. Het doel van deze studie is om de haalbaarheid, consistentie en betrouwbaarheid te onderzoeken van het gebruik van de EQ-5D in de Belgische populatie op basis van een bevraging bij een Belgische subpopulatie.

De vragenlijst werd verdeeld onder 274 gezondheidswerkers. De haalbaarheid van een grootschalig onderzoek werd geëvalueerd aan de hand van het responspercentage, de perceptie van moeilijkheid en de tijd die men nodig had om de vragenlijst in te vullen. Wat consistentie in de waarderingen betreft, werd er gekeken naar de mate waarin gezondheidstoestanden die, wat ernst betreft, een logische relatie hebben tot elkaar, ook als dusdanig worden gewaardeerd. Waarderingen van toestanden die tweemaal voorkomen in de vragenlijst werden vergeleken om de betrouwbaarheid van de vragenlijst na te gaan. Aangezien de gegevens niet normaal verdeeld waren, werden nonparametrische testen gebruikt bij de statistische analyse.

Het responspercentage was 49%. De gemiddelde consistentie was 97,3%. Betrouwbaarheid varieerde van 70 tot 80%. Gegeven het hogere responspercentage en het hogere niveau van consistentie in onze steekproef van gezondheidswerkers in vergelijking met resultaten van populatiestudies in andere landen, kunnen we concluderen dat het zinvol is om het EQ-5D waarderingsinstrument te gebruiken in grootschalig onderzoek in België om sociale preferenties met betrekking tot gezondheidstoestanden af te leiden.

References

1. Gudex C, Dolan P, Kind P, Williams A. Health state valuations from the general public using the visual analogue scale. Qual Life Res 1996; 5: 521-31.
2. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ 1996; 5: 141-54.
3. Williams A. The role of the EuroQol instrument in QALY calculations. York Centre for Health Economics Discussion Paper 1995; 130: 1-13. University of York.
4. Polit DF, Hungler BP. Nursing Research. Principles and Methods. 4th edition. Philadelphia: J.B. Lippincott Company; 1991.
5. The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy 1990; 16: 199-208.
6. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997; 35(11): 1095-108.
7. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37: 53-72.
8. Essink-Bot ML, Stouthard MEA, Bonsel GJ. Generalizability of valuations on health states collected with the EuroQol-questionnaire. Health Econ 1993; 2: 237-46.
9. Harper R, Brazier J, Waterhouse JC, Walters SJ, Jones NMB, Howard P. Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. Thorax 1997; 52: 879-87.
10. Hurst NP, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown M. Validity of EuroQol - a generic health status instrument - in patients with rheumatoid arthritis. B J Rheumatol 1994; 33: 655-62.
11. Dorman P, Slattery J, Farrell B, Dennis M, Sandercock P. Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 questionnaires after stroke. United Kingdom Collaborators in the International Stroke Trial. Stroke 1998; 29: 63-8.
12. Badia X, Monserrat S, Roset M, Herdman M. Feasibility, validity and test-retest reliability of scaling methods for health states: the visual analogue scale and the time trade-off. Qual Life Res 1999; 8: 303-10.
13. Van Agt HME, Essink-Bot ML, Krabbe PFM, Bonsel GJ. Test-retest reliability of health state valuations collected with the EuroQol questionnaire. Social Sci Med 1994; 11: 1537-44.
14. MVH Group. The measurement and valuation of health: first report of the main survey. 1994. York, Centre for Health Economics.
15. Ünal G, De Boer JB, Borsboom GJJM, Brouwer JT, Essink-Bot ML, De Man RA. A psychometric comparison of health-related quality of life measures in chronic liver disease. J Clin Epidemiol 2001; 54: 587-96.
16. Fransen M, Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. Rheumatology 1999; 38(9): 807-13.
17. Stavem K, Bjornaes H, Lossius MI. Properties of the 15D and EQ-5D utility measures in a community sample of people with epilepsy. Epilepsy Res 2001; 44: 179-89.
18. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. Journal of Health Economics 1996; 15: 209-31.

19. Nord E. EuroQol: health-related quality of life measurement. Valuations of health states by the general public in Norway. Health Policy 1991; 18: 25-36.
20. Brooks R, Jendteg S, Lindgren B, Persson U, Björk S. EuroQol: health-related quality of life measurement. Results of the Swedish questionnaire exercise. Health Policy 1991; 18: 37-48.
21. Dolan, P, Kind, P. Inconsistency and health state valuations. Soc Sci Med 1996; 42(4): 609-15.
22. Badia X, Fernandez E, Segura A. Influence of socio-demographic and health status variables on evaluation of health states in a Spanish population. European Journal of Public Health 1995; 5: 87-93.
23. Dolan P. The effect of experience of illness on health state valuations. J Clin Epidemiol 1996; 49(5): 551-64.
24. Kind P, Dolan P. The effect of past and present illness experience on the valuations of health states. Med Care 1995; 33(4 Suppl): AS255-63.
25. Nord E, Richardson J, Macarounas-Kirchman K. Social evaluation of health care versus personal evaluation of health states. Int J Technol Assess Health Care 1993; 9(4): 463-78.
26. Devlin N, Hansen P, Kind P, Williams A. Logical inconsistencies in survey respondents' health state valuations — a methodological challenge for estimating social tariffs. Health Econ 2003; 12(7): 529-44.
27. Devlin N, Williams A. Valuing quality of life: results for New Zealand health professionals. N Z Med J 1999; 112: 68-71.