[Environmental Health and Preventive Medicine 6, 189–191, October 2001]

Original Article

A Study of the Reliability of Health State Valuations in the Japanese EuroQol Instrument

Chisato HAMASHIMA[1] and Katsumi YOSHIDA[1]

[1]*Department of Preventive Medicine St. Marianna University School of Medicine, Kanagawa*

Abstract

Objective: Although the Japanese EuroQol instrument was developed in 1998 by a Japanese translation team, the reliability of health state valuations of the fuller questionnaire has not yet been investigated in Japan. We carried out a pilot study to investigate the reliability of health state valuations of the Japanese EuroQol instrument.

Methods: The survey was conducted twice with the fuller questionnaire of the Japanese EuroQol instrument in a class of preventive medicine. We analyzed test-retest reliability based on the health state valuations and calculated Cronbach's coefficient alpha of both tests.

Results and Conclusions: The mean valuations of 14 hypothetical health states and 'death' between test and retest were not significantly different by paired t-test. The reliability by calculated correlation between test and retest was 0.996 ($p<0.0001$). In addition, Cronbach's coefficient alpha of the first test was 0.827, and that of the second test was 0.865. Although good reliability was shown in the present study, our conclusion was limited to applications of population-based surveys because of the small number and limited subjects of the present study. Further investigations are required in the form of a population-based survey.

Key words: health related QOL, EuroQol, visual analogue scale (VAS), health state valuations, reliability

Introduction

With the steady increase in the prevalence of chronic diseases, health state measurements are increasingly expected to evaluate using indicators of quality of life (QOL). The EuroQol instrument was developed in 1987 by an international research network for describing and evaluating quality of life[1)]. Of particular importance to the EuroQol instrument is the capacity to generate cross-national comparisons of health state valuations and application of economic evaluations[1)]. The fuller questionnaire of the EuroQol includes a section that asks for valuation of hypothetical health states as described in the five dimensional items using the visual analog scale (VAS). The valuation task is intended for use to compare health state valuations held by different population groups, or between different groups within the field of health care. Health state valuations have been conducted in European countries[2–5)] and the United States[5–6)]. Although the Japanese EuroQol instrument was developed in 1998 by a Japanese translation team[7)], the reliability of health state valuations has not yet been investigated in Japan. We conducted a pilot study to investigate the reliability of health state valuations of the Japanese EuroQol instrument.

Received May 14 2001/Accepted Jul. 5 2001
Reprint requests to: Chisato HAMASHIMA
Department of Preventive Medicine St. Marianna University School of Medicine, 2-6-1, Sugao, Miyamae-ku, Kawasaki-shi, Kanagawa
TEL: +81(44)977-8111 (ext. 3427), FAX: +81(44)977-8356

Materials and Methods

The EuroQol instrument

The EuroQol instrument is a generic health state index of health-related QOL on a one-dimensional scale[1)]. A fuller questionnaire of the EuroQol has four components of the instrument: description of the respondent's own health by means of five dimensions (EQ5D), rating of own health using VAS, background information about respondent and valuation task of hypothetical health states defined by five dimensions[3)]. On the other hand, a clinical version of the EuroQol consists of head three compartments, which is most used in clinical research, population health surveys and economic evaluation. The five dimensions of health states are shown; mobility, self-care, usual activities, pain or discomfort, anxiety or depression. The response to each dimension has three categories of the general form 'no problems (level 1)', 'some problems (level 2)', and 'extreme problems (level 3)'. Health state can be described by one level of each dimension. Thus, 243 ($=3^5$) health states, 'death' and 'unconsciousness' included in the EuroQol instrument. The health state described by five dimensions could be changed to a utility score using the tariff that is a table of all possible health states and their corresponding values. These values are numbers on an interval scale with 1 for full health and 0 for health state equivalent to death and these numbers serve as quality adjustment weights to form QALYs

(quality adjusted life years). In the fuller questionnaire, respondents are asked to value 14 hypothetical health states and 'death' as described in the five dimensional items on a VAS with marked end-points: 0='worst imaginable health state' and 100='best imaginable health state'. The health states are fixed in the fuller questionnaire of the standard EuroQol[8].

The subjects

This survey was conducted with the fuller questionnaire of the Japanese EuroQol instrument in a class of preventive medicine. The questionnaires were given out to all students who attended the class with a brief oral explanation about the purpose of this survey and writing method. The first survey was conducted in October, 1998 and one month later, in the same class. A total of 87 medical students participated in the first test, and 85 in the second test. The subjects who participated in both surveys were 65 medical students (31 males and 34 females). The mean age of the 65 subjects was 21.8±1.6 yrs in males and 21.1±1.1 yrs in females.

Statical analysis

We analyzed test-retest reliability based on the findings of the 65 medical students. The differences in their own health valuations between the test and retest were analyzed by the paired t-test. The tariff score was calculated using the Japanese tariff[9]. The mean scores per health state valuation on a VAS were compared between test and retest using the paired t-test. Correlation between the first and the second test was calculated based on the mean score of each health states valuation on the VAS. Cronbach's coefficient alpha was calculated based on a coefficients matrix of health states correlations in both tests.

Results

The health states of the subjects who participated in both surveys are shown in Table 1. Both the tariff and VAS scores of the first test were not significantly different according to the paired t-test. This suggested that the health states of these subjects were equal in both tests. The mean and median valuation on VAS in both the test and retest are shown in Table 2. The mean valuation of 14 hypothetical health states and 'death' were not significantly different between test and retest findings. The VAS score of health state '11111' was higher in the second test than that of the first test (p=0.0215). In addition, the tendency of smaller standard deviations for the descriptions of 'extreme' health states (death and '33333'). This correlation of the mean VAS between both tests was 0.996 (95%CI; 0.988–0.999). Cronbach's coefficient alpha of the first test was 0.827, and that of the second test was 0.865.

Table 1 Subjects of the test-retest reliability

	First test	Second test	P-value
Number	65	65	
Tariff score	0.888±0.148 (0.537–1.000)	0.873±0.145 (0.533–1.000)	0.8914
VAS score	66.8±17.7 (10–95)	64.5±20.9 (10–95)	0.4044

The subjects of the test-retest reliability participated in both tests.
** mean±SD.
(minimum-maximum).

Discussion

Although several studies[2–4] have reported a good reliability of health state valuations in the EuroQol, societal and cultural background factors might affect the results. We conducted a pilot study to investigate the reliability of health state valuations in a fuller questionnaire of the Japanese EuroQol instrument. The York Group studied test-retest reliability as a part of a large valuation study of 3,395 members of the general British population[3]. The respondents were found to be very reliable, with mean intraclass coefficients of 0.78 for VAS. The Rotterdam Health Survey based reported test-retest reliability of health states valuations[2] using the fuller questionnaire of the EuroQol: 0.86 for group level coefficients calculated for each state and averaged over health states and 0.90 for a coefficient derived from individual corrections considering all health states simultaneously. Although Badia et al. reported 0.90 for test-retest reliability, the target group was only 50 persons selected from a population-based study[4]. Lands et al.[10] reported that typical Pearson correlation between test and retest was in the range 0.65 to 0.95, and values above 0.85 might be consid-

Table 2 Health state valuations of test-retest

	First test			Second test			P-value
Health state	Average	SD	Median	Average	SD	Median	
11111	90.6	13.8	95.0	93.3	9.5	100.0	0.0215
21111	64.6	15.0	65.0	60.2	17.3	60.0	0.4630
11112	63.0	17.3	62.0	63.7	15.6	60.0	0.9025
11211	62.4	16.2	60.0	62.8	17.5	65.0	0.8163
11121	61.8	17.7	60.0	58.5	19.2	60.0	0.2442
12111	53.0	18.7	50.5	51.3	16.7	55.0	0.6126
11122	43.1	19.6	40.0	43.3	13.5	40.0	0.3736
21232	31.8	17.3	30.0	34.1	16.8	30.0	0.2550
32211	31.1	18.2	30.0	28.7	18.2	30.0	0.2810
22233	22.9	17.1	20.0	19.9	14.4	20.0	0.3369
33321	18.2	13.9	15.0	14.6	11.9	12.5	0.1068
22323	15.3	12.2	10.0	16.5	14.5	15.0	0.7549
unconscious	13.0	20.6	5.0	14.9	21.0	6.0	0.5571
33333	9.0	11.1	5.0	9.0	13.4	5.0	0.6273
death	8.2	11.9	5.0	7.6	10.9	0.0	0.3220

(n=65)

ered acceptable. The Pearson correlation quantifies the association between two measurement scales, but does not indicate agreement. For measurements made on a continuous scale, the intraclass correlation is increasingly used to indicate reliability instead of Pearson or rank-order coefficients. The test-retest correlation in the present study was 0.996 at the group level ($p<0.0001$) and the confidence interval of the correlation was extremely narrow (95%CI=0.988–0.999). In addition, Spearman's rank correlation between the test and retest was 0.979, which is significant ($p=0.0003$). On the other hand, Cronbach's coefficient alpha of the first test was 0.827, and that of the second test was 0.865, showing a good reliability of health state valuations for the Japanese EuroQol instrument.

Although the correlation of test-retest findings is related to the reliability of the test, test-retest methods have some limitations[11–13]. Nunnally[13] recommended that the interval of the test is between two weeks to one month after the first test. However, he referred to the recollection effect, in which subjects remember their previous answer to the questionnaire. The Rotterdam Health Survey chose an interval test-retest of ten months[2]. On the other hand, in the York Study[3], the interval for the test-retest was an average of ten weeks. In the present study, the test-retest interval was only one month and, therefore, we could not exclude the recollection effect. The present study had an additional restriction, regarding the subjects. Larger studies can tolerate lower standards of reliability measurement than smaller ones, because large sample sizes reduce the error of measurement in estimating mean values. The number of subjects in our study was limited (65 persons), and they shared a similar background. They were all medical students in the same class, with similar ages, education, and place of residence. In addition, a prevalence of moderate or severe problems in the subjects was higher than that of subjects the same age in a population survey[14–15]. The characteristics of the subjects in the present study contained a bias[11–12]. However, a similar study concerning the test-retest reliability of health state valuations in the EuroQol instrument was determined to be good in a Dutch population[2]. Enough participants who have various backgrounds should be chosen for population-based studies. The reliability of the present study might be overestimated because of these reasons. Although good reliability was shown in the present study, our conclusion was limited to be applications of population-based surveys. The questionnaire including hypothetical health state valuations has not been used in a large study in Japan, because the concept to evaluate a hypothetical health state is difficult to understand for some respondents without any explanation[14–16]. The present study was conducted as a pilot study because of several limitations, which were the small number and the characteristics of the subjects. Further investigations of the reliability of health state valuations are required in the form of a population-based survey.

References

1) Brooks R with EuroQol group. EuroQol: the current state of play, Health Policy 1996; 37: 53–72.
2) Van Agt HME, Essink-Bot ML, Krabbe PFM, Bonsel GJ. Test-retest reliability of health state valuations collected with EuroQol questionnaire. Soc. Sci. Med. 1994; 39: 1537–1544.
3) MVH Group. The measurement and valuation of health: First report of main survey. Centra for Health Economics. York. 1994.
4) 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: 303–310.
5) Johnson JA, Ohinmaa A, Murti B, Sintonen H, Coons SJ. Comparison of Finnish and U.S.-based visual analog scale valuations of the EQ-5D measure. Med. Decis. Making. 2000; 20: 281–289.
6) Johnson JA, Ergo A, Coons SJ. Valuation of EuroQol (EQ-5D) in adult US sample. Pharmacoeconomics 1998; 13: 421–433.
7) Japanese EuroQol translation team. The development of the Japanese EuroQol instrument. J. Health Care and Society. 1998; 8(1): 109–123
8) The EuroQol Group. EQ-5D user guide 6th issue. 1999.
9) Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, Hamashima C, Hisashige A, Tamura M. Estimating an EQ-5D population value set: the case of Japan. Health Economics (in press).
10) Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrcs. 1977; 33: 159–174.
11) Uly-del Groot CA, Rutten FFH, Bonsel GJ. Measurement and valuation of quality of life in economic appraisal of cancer treatment. Euro. J. Cancer.1994; 30A: 111–117.
12) Carmines EG, Zeller RA. Reliability and validity assessment. SAGE publications. Bevery Hills. London. 1979.
13) Nunnally JC. Educational measurement and evaluation. New York. McGraw-Hill. Riley MW. 1964.
14) Tsuchiya A, Hasegawa H, Nishimura S, Hisashige A, Ikegami N, Ikeda N. A validity study of the Japanese EuroQol instrument. J. Health Care and Society. 1998; 8(1): 67–76.
15) Ikeda S, Ikegami N. on behalf of the Japanese EuroQol Tariff Project. Health status in Japanese population: results from Japanese EuroQol study. J. Health Care and Society. 1999; 9(3): 83–92.
16) Nawata S, Yamada Y, Ikeda S, Ikegami N. EuroQol study of the elderly general population: relationship with IADL and other attributes. J. Health Care and Society. 2000; 10(2): 75–86.