Health and Quality of Life Outcomes

BioMed Central

Research

Open Access

Do English and Chinese EQ-5D versions demonstrate measurement equivalence? an exploratory study

Nan Luo[1], Ling-Huo Chew[2], Kok-Yong Fong[3,4], Dow-Rhoon Koh[3,4], Swee-Cheng Ng[4], Kam-Hon Yoon[3,4], Sheila Vasoo[4], Shu-Chuen Li[1] and Julian Thumboo*[3,4]

Address: [1]Department of Pharmacy, National University of Singapore, Singapore, [2]School of Health Sciences, Nanyang Polytechnic, Singapore, [3]Department of Medicine, National University of Singapore, Singapore and [4]Department of Medicine, National University Hospital, Singapore

Email: Nan Luo - scip9152@nus.edu.sg; Ling-Huo Chew - CHEW_Ling_Huo@nyp.gov.sg; Kok-Yong Fong - grifky@sgh.com.sg; Dow-Rhoon Koh - phskohdr@nus.edu.sg; Swee-Cheng Ng - Swee_Cheng_Ng@alexhosp.com.sg; Kam-Hon Yoon - mdcykh@singnet.com.sg; Sheila Vasoo - sheilav@nuh.com.sg; Shu-Chuen Li - phalisc@nus.edu.sg; Julian Thumboo* - grijt@sgh.com.sg

* Corresponding author

Published: 17 April 2003

Health and Quality of Life Outcomes 2003, **1**:7

Received: 11 March 2003
Accepted: 17 April 2003

This article is available from: http://www.hqlo.com/content/1/1/7

© 2003 Luo et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Although multiple language versions of health-related quality of life instruments are often used interchangeably in clinical research, the measurement equivalence of these versions (especially using alphabet vs pictogram-based languages) has rarely been assessed. We therefore investigated the measurement equivalence of English and Chinese versions of the EQ-5D, a widely used utility-based outcome instrument.

Methods: In a cross-sectional study, either EQ-5D version was administered to consecutive outpatients with rheumatic diseases. Measurement equivalence of EQ-5D item responses and utility and visual analog scale (EQ-VAS) scores between these versions was assessed using multiple regression models (with and without adjusting for potential confounding variables), by comparing the 95% confidence interval (95%CI) of score differences between these versions with pre-defined equivalence margins. An equivalence margin defined a magnitude of score differences (10% and 5% of entire score ranges for item responses and utility/EQ-VAS scores, respectively) which was felt to be clinically unimportant.

Results: Sixty-six subjects completed the English and 48 subjects the Chinese EQ-5D. The 95%CI of the score differences between these versions overlapped with but did not fall completely within pre-defined equivalence margins for 4 EQ-5D items, utility and EQ-VAS scores. For example, the 95%CI of the adjusted score difference between these EQ-5D versions was -0.14 to +0.03 points for utility scores and -11.6 to +3.3 points for EQ-VAS scores (equivalence margins of -0.05 to +0.05 and -5.0 to +5.0 respectively).

Conclusion: These data provide promising evidence for the measurement equivalence of English and Chinese EQ-5D versions.

Background

With health-related quality of life (HRQoL) being increasingly used as an endpoint in multi-national clinical trials, it is often necessary to use two or more language versions of a HRQoL instrument in any given study. In such studies, it would be ideal to pool data from each language version of a HRQoL instrument for statistical analysis to increase the statistical power and representativeness of such research [1]. In order to do so, these different language versions should measure the same construct (i.e. dimensions of HRQoL) with the same metric; in other words, language versions of a HRQoL instrument should demonstrate measurement equivalence. According to Drasgow and Kanfer [2], different language versions of a HRQoL instrument would demonstrate measurement equivalence if they yielded similar scores at item and scale levels for respondents with identical levels of HRQoL.

Measurement equivalence, also referred to metric equivalence [3], differs from conceptual equivalence and psychometric equivalence. Conceptual equivalence refers to similarity in meaning of items across language versions of an instrument [3], and aims to ensure that different language versions measure the same construct. Psychometric equivalence refers to similarity in psychometric properties such as floor and ceiling effects, reliability and construct validity [4]. Conceptual equivalence and psychometric equivalence are prerequisites for measurement equivalence but do not necessarily ensure measurement equivalence. To date, few studies have investigated measurement equivalence for different language versions of HRQoL instruments [5].

Measurement equivalence of language versions of an instrument is of particular concern in several situations, which are addressed in this study. First, measurement imprecision is more likely to arise in brief instruments. The EQ-5D self-report questionnaire [6] exemplifies this, as information for each of its 5 dimensions is derived from only one item. Second, versions of an instrument in languages which differ in semantic structure may also be more prone to imprecision. For example, score differences are more likely to arise in comparing scores of English and Chinese versions of an instrument (which are alphabet and pictogram-based respectively) than in comparing scores from 2 alphabet-based versions.

The purpose of this study was therefore to evaluate the measurement equivalence of English and Chinese versions of the EQ-5D [6], a brief utility-based HRQoL instrument which is widely used in multi-national clinical trials [7]. The English and Chinese EQ-5D versions used in this study were adapted for use in Singapore using the EuroQol Group's cultural adaptation guidelines [8], thus facilitating conceptual equivalence, and (at the time this manuscript was published) are regarded as 'best available' language versions by the EuroQol Group's Translation Committee [9]. These EQ-5D versions have demonstrated similar psychometric properties [10,11], suggesting psychometric equivalence. Thus, in the current study, we aimed to evaluate the measurement equivalence of these versions by studying if differences in item responses and scale scores between these versions exceeded pre-defined values (corresponding to the minimal clinically important difference) in patients with rheumatic diseases.

Methods

Study design

A consecutive sample of outpatients with rheumatic diseases seen at a tertiary referral hospital within a 2-week period were interviewed by trained nurse interviewers using an identical English or Chinese questionnaire containing the Singaporean English or Chinese EQ-5D, a 10 cm pain visual analog scale (VAS), and assessing psychosocial, socio-demographic and other variables. Written consent was obtained from each subject for this IRB-approved study. Inclusion criteria were physician diagnosis of a rheumatic disease and ability to cooperate with the interview. This research was part of a larger study of English and Chinese EQ-5D versions in subjects with rheumatic diseases [10,11].

Instruments

The *EQ-5D* consists of a health descriptive system and a visual analog scale (EQ-VAS) for respondents to self-classify and rate their health on the day of administration of the instrument [6,7]. The descriptive system has 5 items/dimensions (i.e., mobility, self-care, usual activities, pain/discomfort and anxiety/depression), and for each item, there are three response levels (i.e., no problems, moderate problems and extreme problems). The items can be used individually or in combination (as a health profile) as descriptive measures in clinical studies. Theoretically, the design of the descriptive system identifies as many as 243 unique health states, although a small number of these health states are not plausible in reality [12]. Scoring methods have been developed to assign each of these health states a utility score in which 1 represents full health (no problems with all 5 items) and 0 represents being dead [12,13]. EQ-5D utility scores can be used to calculate quality-adjusted life-years (QALYs) for cost-utility analysis of health interventions [14,15]. The EQ-VAS is a vertical, graduated (0–100 points) 20 cm 'thermometer', with 100 at the top representing 'best imaginable health state' and 0 at the bottom representing 'worst imaginable health state'. The EQ-VAS score can be used as a measure of clinical outcome, using individual respondents' own judgment [16].

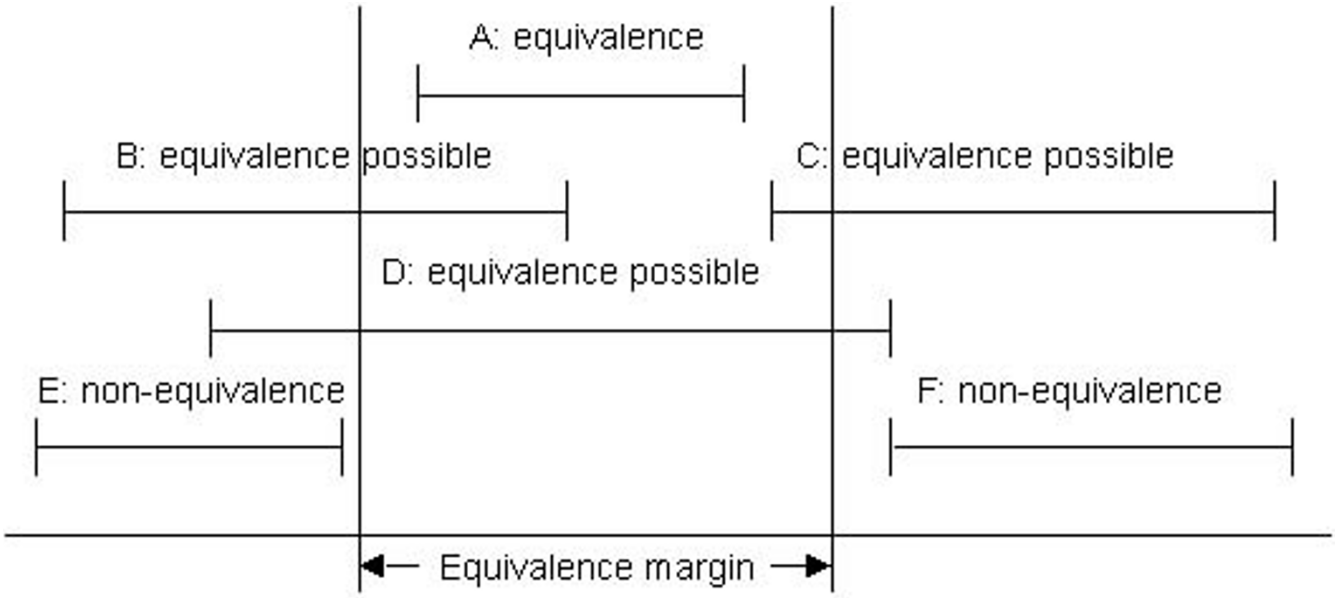

Figure 1
Possible relationships between equivalence margins and 95% confidence intervals (95%CI) of score difference between English and Chinese EQ-5D versions Note: Horizontal lines represent the 95% confidence intervals of difference in EQ-5D scores between English and Chinese versions.

Singaporean English and Chinese EQ-5D versions were derived by adapting the UK English [7] and the Taiwan Chinese versions respectively, using the EuroQol Group's guidelines for cultural adaptation of the EQ-5D [8]. The resulting Singaporean English EQ-5D is identical to its source UK English version except that the word 'box' in the instructions for the EQ-VAS was replaced with 'BLACK BOX' [10]. We found this amendment improved respondents' compliance with EQ-VAS instructions to link the box representing 'your own health state today' to the scale. Changing the word 'box' to 'BLACK BOX' was also adopted for the Singaporean Chinese EQ-5D [11].

Statistical analysis

We evaluated the equivalence of English and Chinese EQ-5D versions by examining whether score differences between these versions were clinically important. Based on the definition of Drasgow and Kanfer [2], if these differences were clinically unimportant, these versions would demonstrate measurement equivalence. Using methodology for assessing therapeutic equivalence in clinical trials [17,18], we therefore compared the 95% confidence interval (95%CI) of EQ-5D item, utility and VAS score differences with pre-defined equivalence margins to determine if differences in scores were clinically important or unimportant. Each equivalence margin represented a range of score differences which would be too small to be clinically important. Comparing the 95% CI for a score difference with its corresponding equivalence margin could lead to 1 of 3 possible results [17,18], illustrated graphically in Figure 1. First, if the 95%CI fell completely within the equivalence margin, the score difference would be clinically unimportant, and measurement equivalence would be demonstrated (Figure 1, option A). Second, if the 95%CI did not overlap with the equivalence margin at all, the score difference would be clinically important, and non-equivalence would be demonstrated (Figure 1, options E and F). Third, if the 95%CI partially overlapped with the equivalence margin, the score difference might or might not be clinically important, and either equivalence or non-equivalence might be present (Figure 1, options B, C and D).

We defined equivalence margins for responses to items, utility and EQ-VAS scores after calculating utility scores using the algorithm developed by Dolan [12]. Ideally, each equivalence margin should be based on the minimal clinically important difference (MCID) [19] for that score. However, as these have not been specified in the EQ-5D literature, we defined equivalence margins based on the best available data (see below). Score differences were assessed using logistic regression models for responses to

Table 1: Characteristics of subjects completing the English or Chinese EQ-5D

	n (%) unless stated		
	English (n = 66)	Chinese (n = 48)	*p*-value*
Mean ± SD (median) age	44.3 ± 17.2 (43.0)	56.7 ± 12.4 (57.0)	<0.001
Female gender	48 (72.7)	45 (93.8)	0.006
Chinese ethnicity	45 (68.2)	48 (100)	-
Employed/full time student	38 (57.5)	13 (27.1)	0.010
Six or less years of education	8 (12.1)	31 (64.6)	<0.001
Married	40 (60.6)	38 (79.2)	0.597
Acute medical condition present†‡	50 (75.8)	36 (75.0)	0.926
Chronic medical condition present§	27 (40.9)	28 (58.3)	0.066
Mean ± SD (median) 10 cm pain VAS score‡	3.3 ± 2.7 (2.6)	4.7 ± 2.2 (4.6)	0.003
Tender points present	42 (63.6)	31 (64.6)	0.883
Rheumatic disease			
Osteoarthritis	9 (13.6)	15 (31.3)	0.006
Rheumatoid Arthritis	26 (39.4)	23 (47.9)	
Systemic Lupus Erythematosus	23 (34.8)	8 (16.7)	
Spondyloarthropathy	8 (12.2)	2 (4.1)	
Response to EQ-5D items			
Mobility			
No problems	50 (75.8)	40 (83.3)	0.327
Some problems	16 (24.2)	8 (16.7)	
Extreme problems	0 (0)	0 (0)	
Self-care			
No problems	66 (100)	46 (95.8)	0.094
Some problems	0 (0)	2 (4.2)	
Extreme problems	0 (0)	0 (0)	
Usual activities			
No problems	50 (75.8)	43 (89.6)	0.125
Some problems	15 (22.7)	4 (8.3)	
Extreme problems	1 (1.5)	1 (2.1)	
Pain/discomfort			
No pain/discomfort	14 (21.2)	11 (22.9)	0.889
Moderate pain/discomfort	49 (74.2)	34 (70.8)	
Extreme pain/discomfort	3 (4.5)	3 (6.3)	
Anxiety/depression			
No anxiety/depression	40 (60.6)	30 (62.5)	0.098
Moderate anxiety/depression	26 (39.4)	15 (31.2)	
Extreme anxiety/depression	0 (0)	3 (6.3)	

*Chi-square or *t*-test † Acute medical conditions included upper respiratory tract infections, vomiting or diarrhoea, headache lasting more than 1 day, insomnia and injuries. ‡ The recall period for the pain VAS and acute medical conditions was the preceding 4 weeks. §Chronic medical conditions included hypertension, diabetes mellitus, stroke, cancer, joint replacement and limb fractures.

EQ-5D items and linear regression models for EQ-5D utility and EQ-VAS scores, with or without adjustment for influence of other variables which might influence HR-QoL (e.g. age, gender, socio-economic status). It was important to adjust for the influence of these variables when assessing influence of language, as observed differences in EQ-5D scores might be caused by these determinants rather than by questionnaire language.

Score differences for each EQ-5D item were examined using logistic regression models (one model for each item). Responses to each item were treated as a binary dependent variable (no problems = 0/with problems = 1) by combining response levels 2 (moderate problems) and 3 (extreme problems). Language version was coded into a dummy variable (Chinese = 0/English = 1). Each model was constructed by first entering language version as the only independent variable; other selected independent variables were entered subsequently. These variables were age, gender, years of education, employment status and pain VAS score, which were selected because they differed substantially between the two groups (Table 1) and have been reported to correlate with responses to the EQ-5D [20–26]. Age and pain VAS scores were treated as continuous

Table 2: Logistic regression: the influence of language version on EQ-5D item responses

Dependent variable	Unadjusted influence of language version			Adjusted influence of language version*		
	Odds ratio (95%CI)	*p*-value	Corresponding proportion interval	Odds ratio (95%CI)	*p*-value	Corresponding proportion interval
Mobility	1.60 (0.62, 4.12)	0.330	(-17.0% to +9.8%)	2.09 (0.60, 7.30)	0.249	(-20.0% to +10.5%)
Usual activities	2.75 (0.93, 8.13)	0.067	(-20.4% to +1.4%)	4.16 (1.01, 17.09)	0.048	(-22.4% to -0.2%)
Pain/discomfort	1.10 (0.45, 2.70)	0.828	(-20.9% to +10.4%)	1.21 (0.38, 3.87)	0.752	(-29.9% to +11.9%)
Anxiety/depression	1.08 (0.50, 2.33)	0.838	(-17.6% to +17.1%)	0.97 (0.34, 2.77)	0.956	(-20.4% to +26.3%)

95%CI = 95% confidence interval; Chinese language was the reference group; self-care was excluded because no problems were reported by subjects completing the Singaporean English EQ-5D. *Adjusted for the influence of age, gender, years of education, employment status and pain VAS score. Pain VAS score was not included in the model for the pain/discomfort item.

variables while the remaining variables were coded into dummy variables. In the model for pain/discomfort item, pain VAS score was not included, as both variables measure pain and therefore did not represent an independent-dependent relationship (a necessary assumption for regression models [27]). Based on previous study designs for detecting therapeutic equivalence [18,28], we pre-defined an equivalence margin of (-10%, +10%) for each EQ-5D item. Thus, if the difference in the proportion of subjects reporting problems for an item was less than ± 10% between the 2 language groups, the item demonstrated measurement equivalence as defined for this study. The 95%CI of the regression coefficient (i.e. odds ratio) for language version was converted into the 95% CI of the proportion derived from this odds ratio to facilitate comparison with the equivalence margin of (-10%, +10%) [29].

Score differences in EQ-5D utility and EQ-VAS scores were examined using separate linear regression models, with and without adjustment for variables potentially influencing HRQoL (listed above). The 95%CI of the regression coefficient of language version was compared with pre-defined equivalence margins to determine measurement equivalence. We reviewed the literature but found no reports regarding the MCID of EQ-5D utility or EQ-VAS scores; however, we found that differences in mean EQ-5D utility and EQ-VAS scores in individuals differing in health status were seldom less than 5% of their entire score ranges (i.e., 0.05 points for utility scores and 5.0 points for EQ-VAS scores) [23–25,30–33]. We therefore defined equivalence as a difference of less than 5% of the respective score ranges, that is, (-0.05, +0.05) for utility and (-5.0, +5.0) for EQ-VAS scores.

Results

Subject characteristics

The mean (SD) age of the 114 interviewed subjects was 49.4 (16.5) years, with 81.6% (n = 93) being female and 57.9% (n = 66) completing the English EQ-5D. English EQ-5D respondents were younger (44.3 vs. 56.7 years, $p < 0.001$) and more likely to be male (27.3% vs. 6.2%, $p < 0.01$), employed (57.5% vs. 27.1%, $p = 0.01$), better educated (6 or less years of education: 12.1% vs. 64.6%, $p < 0.001$) and reported less pain (mean pain VAS score: 3.3 vs. 4.7, $p < 0.01$) than Chinese-speaking respondents. Other characteristics of the two groups of subjects are summarized in Table 1.

Comparison of responses to EQ-5D items

Equivalence was assessed for mobility, usual activities, pain/discomfort items and anxiety/depression items, but not for the self-care item as subjects completing the English EQ-5D reported no problems with this item (Table 1), which did not allow a logistic regression model to be created. The 95%CIs of odds ratios and corresponding proportion intervals for language version are summarized in Table 2. Before adjusting for other variables, the influence of language partially overlapped with the equivalence margin of (-10%, +10%) for all items, corresponding to options B or D in Figure 1. The lower bound of the 95%CIs ranged from -20.9 to -17.0%. The upper bound of the 95%CI for mobility, usual activities and pain/discomfort items was less than or approximately +10%; that for the anxiety/depression item was 17.1%. After adjusting for the influence of other variables, the 95%CIs changed slightly but still overlapped partially with the equivalence margin for all items. The lower bound ranged from -29.9 to -20.0%; the upper bound was less than or approximately +10% for 3 items and 26.3% for the anxiety/depression item.

Comparison of EQ-5D utility and EQ-VAS scores

The 95%CIs of effect sizes for language version on EQ-5D utility and VAS scores are summarized in Table 3. Before adjusting for other variables, the influence of language partially overlapped with the equivalence margin for both utility and VAS scores, with the 95%CI of the effect of lan-

Table 3: Linear regression: the influence of language version on EQ-5D utility and visual analog scale scores

Dependent variable	Unadjusted effect size (95%CI)	*p*-value	Adjusted effect size (95%CI)*	*p*-value
EQ-5D utility score	0.01 (-0.07, 0.09)	0.851	-0.05 (-0.14, 0.03)	0.214
EQ-VAS score	2.7 (-3.5, 8.8)	0.396	-4.1 (-11.6, 3.3)	0.276

95%CI = 95% confidence interval; Chinese language was the reference group; EQ-VAS = EQ-5D Visual Analog Scale *Adjusted for the influence of age, gender, years of education, employment status and pain VAS score.

guage version being (-0.07, +0.09) and (-3.5, +8.8), respectively. This corresponded to options D and C respectively in Figure 1. After adjusting for the influence of other variables, the 95%CIs were (-0.14, +0.03) and (-11.6, +3.3) respectively, which again partially overlapped with the respective pre-defined equivalence margins of (-0.05, +0.05) and (-5.0, +5.0).

Discussion

Many HRQoL instruments have been developed in one language for use in one country, then translated and culturally adapted for use in other countries. This approach has been adopted by the International Quality of Life Assessment Project (IQOLA) [34] and the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Study Group [35]. In addition to conceptual and psychometric equivalence, measurement/metric equivalence of different language versions of a HRQoL instrument developed using this approach also needs to be assessed as a necessary prelude to pooling data from these versions into a single analytical framework in clinical trials.

In this study, we investigated the measurement equivalence of Singaporean English and Chinese versions of EQ-5D by applying methodology used to assess therapeutic equivalence of medical interventions in clinical trials. This involved comparing the 95%CI of the score differences between these language versions against a corresponding pre-defined equivalence margin (which corresponded to a magnitude of score differences which were felt to be clinically unimportant). The 95%CI of differences in EQ-5D item responses and utility and VAS scores between these versions, with or without adjustment for confounding variables, partially overlapped with their respective pre-defined equivalence margins. Our data thus provide promising evidence for the equivalence of Singaporean English and Chinese EQ-5D versions, and justify a larger study to conclusively address this issue, possibly matching respondents by health status and socio-demographic characteristics to reduce the potential confounding effects of these factors. Our study is one of few investigations into the measurement equivalence of different language versions of EQ-5D using outcome scores of the EQ-5D. Such studies are meaningful and useful for the various language versions of both the EQ-5D and other HRQoL scales. In a previous study, using item-response theory (IRT) [36], investigators confirmed the cross-cultural comparability of EQ-5D items across 10 different European language versions in outpatients with schizophrenia [37]. These results, though encouraging, cannot be generalized to other language versions of EQ-5D or to subjects without schizophrenia.

Defining an equivalence margin for different language versions of a HRQoL instrument involves specifying the magnitude of difference in scores between versions that is clinically unimportant (i.e. that would not adversely influence the results of research which pooled data from these versions [18]). Theoretically, the equivalence margin should be no larger than the minimal clinically important difference (MCID) of the scales in that instrument. However, to the best of our knowledge, the MCID for the EQ-5D has not been established. The equivalence margins used in this study were therefore estimated from the best available literature, and need to be confirmed in studies specifically assessing the MCID for EQ-5D item responses and utility and EQ-VAS scores.

We recognize several limitations of this study. First, the sample size was relatively small, because data from a validation study were used. In general, studies to assess equivalence need larger sample size [17]. Second, studying subjects attending a rheumatology clinic in a tertiary-care hospital limited the generalizability of our results. However, the results of this exploratory study are encouraging and do justify a larger study to further address the important issue of measurement equivalence for this widely used instrument.

Conclusion

In conclusion, the results of this exploratory study suggest that Singaporean English and Chinese EQ-5D versions may demonstrate measurement equivalence. This study provides justification for further research to investigate the measurement equivalence of these and other EQ-5D language versions, so that results from these versions in clinical trials may be pooled for analysis, thus increasing the representativeness and power of such studies.

Authors' contributions

NL and JT designed and supervised the study, analyzed and interpreted data, and drafted the manuscript. LHC provided administrative and technical support and supervised the study. SCL provided administrative and technical support. All authors except SCL contributed to data collection. All authors made critical revisions of the manuscript for important intellectual content and approved the final manuscript.

Acknowledgements

We would like to thank staff nurses from the Nanyang Polytechnic Advanced Diploma in Nursing Course for their help in interviewing subjects.

References

1. Thumboo J, Fong KY, Chan SP, Machin D, Feng PH, Thio ST and Boey ML **The equivalence of English and Chinese SF-36 versions in bilingual Singapore Chinese** *Qual Life Res* 2002, **11:**495-503
2. Drasgow F and Kanfer R **Equivalence of psychological measurement in heterogeneous populations** *J Appl Psychol* 1985, **70:**662-680
3. Anderson RT, Aaronson NK, Leplege AP and Wilkin D **International use and application of generic health-related quality of life instruments** *In: Quality of life and Pharmacoeconomics in Clinical Trials (Edited by: Spilker B) Philadelphia, Lippincott-Raven* 1996, 613-632
4. Lohr K, Skillman S and with 1999 Health Outcomes Methodology Symposium **Glossary for health outcomes methodology** *Med Care* 2000, **38(suppl):**II7-13
5. Anderson RT, Aaronson NK, Bullinger M and McBee WL **A review of the progress towards developing health-related quality-of-life instruments for international clinical studies and outcomes research** *Pharmacoeconomics* 1996, **10:**336-355
6. Brooks R and with the EuroQol Group **EuroQol: the current state of play** *Health Policy* 1996, **37:**53-72
7. Rabin R and de Charro F **EQ-5D: a measure of health status from the EuroQol Group** *Ann Med* 2001, **33:**337-343
8. The EuroQol Group **Draft guidelines for cultural adaptations of EQ-5D** *Rotterdam* 2000,
9. **EQ-5D available language versions** [http://www.euroqol.org/translations/translations_available2.htm] January 2003
10. Luo N, Chew LH, Fong KY, Koh DR, Ng SC, Yoon KH, Vasoo S, Li SC and Thumboo J **Validity and reliability of the EQ-5D self-reported questionnaire in English-speaking Asian patients with rheumatic diseases in Singapore** *Qual Life Res* 2003, **12:**87-92
11. Luo N, Chew LH, Fong KY, Koh DR, Ng SC, Yoon KH, Vasoo S, Li SC and Thumboo J **Validity and reliability of the EQ-5D self-reported questionnaire in English-speaking patients with rheumatic diseases in Singapore** *Ann Acad Med Singapore*
12. Dolan P **Modelling valuations for EuroQol health states** *Med Care* 1997, **35:**1095-1108
13. Dolan P and Roberts J **Modelling valuations for Eq-5d health states: an alternative model using differences in valuations** *Med Care* 2002, **40:**442-446
14. Gold MR, Patrick DL, Torrance GW, Fryback DG, Hadorn DC, Kamlet MS, Daniels N and Weinstein MC **Identifying and valuing outcomes** *In: Cost-Effectiveness in Health and Medicine (Edited by: Gold MR, Siegel JE, Russell LB, Weinstein MC)* 1996, 82-134
15. Drummond MF, O'Brien BJ, Stoddart GL and Torrance GW **Cost-utility analysis** *In: Methods for the Economic Evaluation of Health Care Programmes Oxford, Oxford University Press* 1997, 139-199
16. Kind P, Hardman G and Macran S **UK population norms for EQ-5D** *York Centre for Health Economics Discussion Paper* 1999, 172
17. Jones B, Jarvis P, Lewis JA and Ebbutt AF **Trials to assess equivalence: the importance of rigorous methods** *BMJ* 1996, **313:**36-39
18. Chadwick D **Monotherapy comparative trials: equivalence and differences in clinical trials** *Epilepsy Res* 2001, **45:**101-103
19. Jaeschke R, Singer J and Guyatt GH **Measurement of health status. Ascertaining the minimal clinically important difference** *Control Clin Trials* 1989, **10:**407-415
20. Brazier J, Jones N and Kind P **Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire** *Qual Life Res* 1993, **2:**169-180
21. Kind P, Dolan P, Gudex C and Williams A **Variations in population health status: results from a United Kingdom national questionnaire survey** *BMJ* 1998, **316:**736-741
22. Johnson JA and Coons SJ **Comparison of the EQ-5D and SF-12 in an adult US sample** *Qual Life Res* 1998, **7:**155-166
23. Badia X, Schiaffino A, Alonso J and Herdman M **Using the EuroQol 5-D in the Catalan general population: feasibility and construct validity** *Qual Life Res* 1998, **7:**311-322
24. Burstrom K, Johannesson M and Diderichsen F **Swedish population health-related quality of life results using the EQ-5D** *Qual Life Res* 2001, **10:**621-635
25. Hurst NP, Kind P, Ruta D, Hunter M and Stubbings A **Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D)** *Br J Rheumatol* 1997, **36:**551-559
26. Wolfe F and Hawley DJ **Measurement of the quality of life in rheumatic disorders using the EuroQol** *Br J Rheumatol* 1997, **36:**786-793
27. Zar JH **Simple linear regression** *In: Biostatistical Analysis New Jersey, Prentice-Hall* 1999, 324-359
28. Rohmel J **Therapeutic equivalence investigations: statistical considerations** *Stat Med* 1998, **17:**1703-1714
29. Bland JM and Altman DG **Statistics notes. The odds ratio** *BMJ* 2000, **320:**1468
30. Statistics Canada **A head-to-head comparison of two generic health status measures in the household population: McMaster Health Utilities Index (Mark 3) and the EQ-5D (internal documents)** *Ottawa* 2000,
31. Hawthorne G, Richardson J and Day NA **A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments** *Ann Med* 2001, **33:**358-370
32. Brazier JE, Harper R, Munro J, Walters SJ and Snaith ML **Generic and condition-specific outcome measures for people with osteoarthritis of the knee** *Rheumatology (Oxford)* 1999, **38:**870-877
33. Wu AW, Jacobson KL, Frick KD, Clark R, Revicki DA, Freedberg KA, Scott-Lennox J and Feinberg J **Validity and responsiveness of the Euroqol as a measure of health-related quality of life in people enrolled in an AIDS clinical trial** *Qual Life Res* 2002, **11:**273-282
34. Aaronson NK, Acquadro C, Alonso J, Apolone G, Bucquet D, Bullinger M, Bungay K, Fukuhara S, Gandek B and Keller S **International Quality of Life Assessment (IQOLA) Project** *Qual Life Res* 1992, **1:**349-351
35. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB and de Haes JC **The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology** *J Natl Cancer Inst* 1993, **85:**365-376
36. Hays RD, Morales LS and Reise SP **Item response theory and health outcomes measurement in the 21st century** *Med Care* 2000, **38(Suppl 9):**II28-42
37. Prieto L, Hormaechea JA, Sacristan JA, Novick D, Edgell ET and Alonso J **Rasch model analysis to test the cross-cultural validity of the EuroQol-5D (EQ-5D) in the schizophrenia outpatient health outcomes (SOHO) study [abstract]** *Qual Life Res* 2002, **11:**658