Original article

Comparison of three utility measures in stroke patients using item response theory analysis

Ryota Izumi[1], Shinichi Noto[1], Shunya Ikeda[2], Takashi Fukuda[3], Takeru Shiroiwa[3], Ataru Igarashi[4]

[1]Department of Occupational Therapy, Faculty of Medical Technology, Niigata University of Health and Welfare, Niigata, Japan
[2]Department of Pharmaceutical Sciences, School of Pharmacy, International University of Health and Welfare, Tochigi, Japan
[3]Center for Public Health Informatics, National Institute of Public Health, Saitama, Japan
[4]Department of Drug Policy and Management, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan

Keywords: utility measures, item response theory, stroke, EQ-5D, HUI3

Received: 16 October 2013/Accepted: 20 February 2014

Abstract

Our study aimed to clarify the measurement characteristics of three utility measures by item response theory (IRT) analysis in stroke patients.

A multicenter cross-sectional study was conducted on patients with stroke. The Japanese versions of the EQ-5D-3L, EQ-5D-5L and HUI3 were used for utility measurement. Two-parameter logistic models were used for the model of IRT analysis. The choice probability (item information) of patients when a patient's capability value theta (health states) is given is expressed with this model. Moreover, discrimination and difficulty, which are the parameters showing the item characteristic, can be shown.

Subjects were 526 stroke patients. The maximum of the total information was 13.1 in the EQ-5D-3L, 22.5 in the EQ-5D-5L and 8.7 in the HUI3. Moreover, of the three measures, the information in the EQ-5D-5L was the highest, and information could be acquired over a wide range of health states ($-1.4 < \text{theta} < 1.1$, information >10.0). Discrimination was good for all of the measures, and the EQ-5D-5L especially had high discrimination. The three utility measures showed a wide range for difficulty. However, theta was not adapted when extremely high or low.

These findings indicate that the EQ-5D-3L, EQ-5D-5L and HUI3 could all measure the stroke patient's utility scores over a wide range of health states. However, in the stroke patients, the EQ-5D-5L gave the most information and was the highest of the three utility measures in terms of discrimination.

Introduction

Economic evaluation of medical technology has been carried out around the worldby public authorities such as the National Institute for Health and Clinical Excellence (NICE) in the UK. The need for economic evaluation of health technologies such as drugs and devices is now finally being discussed in Japan [1]. Quality-adjusted life years (QALYs), a measure that is already used abroad, is being considered for use as an indicator of effect in evaluations of cost-effectiveness that may be used in considering the introduction new products in Japan. QALYs can

Corresponding author: Ryota Izumi, MS
Department of Occupational Therapy, Faculty of Medical Technology, Niigata University of Health and Welfare, 1398 Shimami-cho, Kita-ku, Niigata 950-3198, Japan
TEL/FAX: +81-25-257-4437, E-mail: izumi@nuhw.ac.jp

be used as an important indicator of effect in the field of rehabilitation.

Utility measures for the health-related quality of life (HRQL) index are used to calculate QALYs. However, Japanese versions are available for only three measures: the EuroQol 5 Dimension 3 Level (EQ-5D-3L) [2], the EuroQol 5 Dimension 5 Level (EQ-5D-5L) [3], and the Health Utilities Index Mark 3 (HUI3) [4]. In Japan, accumulated data for utility measures is insufficient, and research to verify measurement characteristics such as reliability and validity are also limited. On this basis, we previously investigated the reliability, validity and responsiveness of the Japanese versions of the EQ-5D-3L and HUI3 measures. Our results suggested that both measures have good reliability, validity and responsiveness [5-6]. Moreover, we showed high-value inter-rater reliability between patients and rehabilitation staff for both measures. However, we felt that verification of measurement characteristics such as discrimination and difficulty against different questions on each scale is also needed. These characteristics can be verified by using item response theory (IRT), but few studies have analyzed utility measures and HRQL scales using IRT.

IRT can be used to select the most useful items for a shortened measure and to develop a scoring algorithm that predicts the total score on the full measure [7,8]. Sasagawa et al [9] pointed out two advantages to applying IRT. First, the measurement accuracy itself can be improved. It is possible that by increasing the measurement accuracy and decreasing misclassification rates at screening, improvements in the efficiency of work can be observed. Moreover, with the ability to accurately evaluate the measured variable to a finer degree, it is possible to accurately record temporal changes for each of the subjects in the study setting when considering the therapeutic effects in the clinical setting. The second advantage is that for subjects with relatively high characteristic values (meaning those with high utility measures in this study) for which measurements were standardized, it is possible to evaluate the degree of measurement accuracy for subjects with low characteristic values. In the present study, because existing scales are used, indication of the latter type is possible.

In terms of previous research showing the measurement characteristics for existing measures, Tokuda et al [10] conducted an IRT analysis using the Japanese version of the Short Form-8 for healthy persons and reported its suitability for subjects with low QOL measures. Fryback et al [11] conducted an IRT analysis of the EQ-5D-3L, HUI2, HUI3, the Quality of Well-Being Index Self-Administered Version (QWB-SA) and the Short Form-6 Dimension (SF-6D) for subjects with various health states. The QWB-SA and the SF-6D were shown to be suitable for subjects with average health states and the EQ-5D-3L, HUI2 and HUI3 were suitable for subjects with poor health. Furthermore, the HUI3 showed the highest degree of discrimination among the 5 measures.

However, examination of the utility measurement in stroke patients has not been considered in Japan and the measurement characteristics for the Japanese version of the HUI3 and the EQ-5D have not been clarified. In the future, when using QALYs that are based on utility measures, consideration of the characteristics of the measurement is thought to be important. Therefore, the purpose of the present study was to clarify the measurement characteristics of the Japanese versions of the HUI3 and the EQ-5D utility measures by IRT analysis.

Methods

1. Subjects and study design

This was a cross-sectional multicenter study. Eight hospitals were recruited from 6 prefectures

(Niigata, Ishikawa, Saitama, Tokyo, Shizuoka and Hyogo) throughout Japan. The study period was from March 25, 2010 to July 24, 2010. The subjects were stroke patients (with cerebral infarction, cerebral hemorrhage, and subarachnoid hemorrhage) undergoing recovery-phase rehabilitation during hospitalization.

This study was conducted based on the "Ethical Guidelines for Epidemiological Research" [12]. Moreover, prior approval was obtained from the Research Ethics Committee of Niigata University of Health and Welfare, and informed consent from each patient and family was obtained before participation in the study.

2. Investigative method

We used the Japanese versions of the EQ-5D-3L, EQ-5D-5L and HUI3 for utility measurement. We used the Barthel Index (BI) for measuring activities of daily living (ADL) and the Modified Rankin Scale (MRS) to determine the degree of dysfunction. We also instructed rehabilitation staff at each center on how to collect the data, as needed.

3. Measurements

1) EQ-5D

The EQ-5D questionnaire is a generic instrument to evaluate health developed by the EuroQol Group [13]. The EQ-5D defines health according to five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The current descriptive system is made up of these five dimensions, each with three levels (EQ-5D-3L) or five levels (EQ-5D-5L). The range of utility scores for the Japanese versions of the EQ-5D-3L and the EQ-5D-5L are 1.00 to -0.111 [14]. The EQ-5D-5L was also calculated with indirect interim mapping methods presented by the EuroQol group [15]. The EQ-5D-3L was calculated using the valuation set from the Japanese EuroQol Translation team [2].

2) HUI3

The HUI is family of generic health profiles and a preference-based system for the purposes of measuring health status, reporting HRQL and producing utility scores. The HUI3 has specified four to six levels per attribute (vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain) and describes 972,000 unique health status markers [16]. Moreover, it is possible to measure simultaneously the Single score and Global score with the HUI3. The range of utility scores for the Japanese version of the HUI3 is 1.00 to -0.36. The HUI3 was calculated with the scoring formula used by Feeny et al in the Canadian version [17].

3) BI

The BI is a measure developed by Mahoney and Barthel that allows for the evaluation of 10 basic items related to ADL, and the score is represented by 0 to 100 points. Each item is weighted, and a number from 0 to 15 points is allotted [18].

4) MRS

The MRS is an evaluation method that is divided into seven levels based on degree of disability [19]: “No symptoms at all (Grade 0)”, “No significant disability despite symptoms (Grade 1)”, “Slight disability (Grade 2)”, “Moderate disability (Grade 3)”, “Moderately severe disability (Grade 4)”, “Severe disability (Grade 5)”, and “Dead (Grade 6)”. Because Grade 6 is not covered in this study, only Grades 0 to 5 were considered.

4. IRT analysis

IRT is a new testing theory to replace classical testing theories, which are referred to as conventional psychometric methods, and has been proposed as a theory to overcome some difficulties inherent in classical testing theory. IRT is rich in potential applications. In addition to measuring achievement abilities, application of IRT to psychometric evaluations of surveys such as personality measures has been proposed [20]. IRT refers to a set of mathematical models that describe, in probabilistic terms, the relation between a person’s response to a survey question and his or her level of the latent variable being measured by the measurement. This latent variable is usually a hypothetical construct, trait, domain or ability, which is postulated to exist but cannot be directly measured by a single observable variable or item. Instead, it is indirectly measured using multiple items or questions in a multi-item measurement. The underlying latent variable, expressed mathematically by theta (θ), may be any measurable construct, such as mental health, fatigue or physical functioning [21].

It is also possible to express the "discrimination" and "difficulty" of questions that have been used in the evaluation measurement based on the pattern of answers from the subjects [22]. Discrimination indicates the strength of the relation between an item and the measured construct. This parameter also indicates how well an item discriminates between respondents below and above the item difficulty, as indicated by the slope of the item characteristic curve. Difficulty indicates the severity or difficulty of an item. Each difficulty is described by between-category threshold parameters. Difficulty represents the trait level necessary to respond above a threshold probability of 0.50. Also, a negative number for difficulty represents an easy category (possible to select if the health state is low), whereas a positive number represents a difficult category (not possible to select if the health state is low). The location is along the θ-continuum of the item response categories [21].

IRT identifies those items that are best for distinguishing subjects with very high functioning from those with slightly less high functioning ("difficult" items) or those items that are more suited to discriminating between subjects with low functioning from those with very low functioning ("easy" items).

By conducting the analysis, the test characteristic curve and total information curve are obtained. The vertical axis of the test characteristic curve shows the total score (Characteristic), and the slope of the curve indicates the item discrimination. The horizontal axis shows the difficulty (Ability of subject), which indicates the mean health states (referred to as the population mean θ = 0). The total information curve represents the total information (Information) in the vertical axis and difficulty in the horizontal axis. The amount of information in the vertical axis is an indicator of the measurement accuracy and is a reliability coefficient in classical test theory. Therefore, a higher amount of information indicates higher measurement accuracy.

The IRT model used was a graded response model, which is an extension of the two-parameter logistic model. The Fisher information in the maximum likelihood estimation was used to analyze the amount of information. The IRT analyses were carried out using IRTPRO 2.1, and IBM SPSS Statistics Version 19.0 was used for all of the other statistical analyses.

Results

1. Characteristics of the subjects and utility scores (Table 1)

The subjects included 526 patients: 321 (61.0%) were men and 205 (39.0%) were women. The mean age was 67.1 years, and the mean period from stroke onset was 79.6 days. The diagnoses included cerebral infarction (n = 289, 54.9%), cerebral hemorrhage (n = 184, 35.0%), subarachnoid hemorrhage (n = 40, 7.6%), and others (n = 13, 2.5%). For the MRS, 3 patients were Grade 0, 64 were Grade 1, 100 were Grade 2, 97 were Grade 3, 216 were Grade 4, and 46 were Grade 5. Mean utility scores were 0.19 for the HUI3, 0.55 for the EQ-5D-3L and 0.52 for the EQ-5D-5L, and the BI was 65.9 points. Single scores for the HUI3 were 0.89 for vision, 0.90 forhearing, 0.81 forspeech, 0.45 forambulation, 0.64 fordexterity, 0.74 foremotion, 0.64 forcognition and 0.84 for pain.

Table 1. Characteristics of the subjects and utility scores

		n (526)	%
Sex (n)	Male	321	60.9
	Female	205	38.9
Diagnosis	Infarction	289	54.8
	Hemorrhage	184	34.9
	SAH	40	7.6
	Other	13	2.5
Modified Rankin Scale	0	3	0.6
	1	64	12.1
	2	100	19.0
	3	97	18.4
	4	216	41.0
	5	46	8.7
		Mean (SD)	95% CI
Age (years)		67.1 (13.0)	66.0-68.2
Period from onset (days)		79.6 (50.1)	75.3-83.8
EQ-5D-3L		0.55 (0.25)	0.53-0.57
EQ-5D-5L		0.52 (0.26)	0.49-0.54
HUI3	Global score	0.19 (0.34)	0.16-0.21
	Vision	0.89 (0.19)	0.87-0.90
	Hearing	0.90 (0.26)	0.87-0.92
	Speech	0.81 (0.26)	0.79-0.83
	Ambulation	0.45 (0.36)	0.42-0.48
	Dexterity	0.64 (0.33)	0.61-0.67
	Emotion	0.74 (0.21)	0.72-0.76
	Cognition	0.64 (0.30)	0.61-0.66
	Pain	0.84 (0.20)	0.82-0.86
Barthel Index		65.9 (31.1)	63.2-68.5

Abbreviations: SAH (subarachnoid hemorrhage), EQ-5D-3L (EuroQol5Dimension 3level), EQ-5D-5L (EuroQol5 Dimension 5level), HUI3 (Health Utilities Index Mark 3), SD (standard deviation), CI (confidence interval).

2. IRT analysis

(1) Discrimination and difficulty (Table 2, Figure 1 (a)–(e))

Discrimination for the EQ-5D-3L was 4.58 for mobility, 4.33 for self-care, 3.25 for usual activities, 1.24 for pain/discomfort and 0.97 for anxiety/depression. That for the EQ-5D-5L was 4.67 for mobility, 6.69 for self-care, 3.39 for usual activities, 1.29 for pain/discomfort and 0.96 for anxiety/depression. That for the HUI3 was 0.51 for vision, 0.70 for hearing, 1.16 for speech, 3.91 for ambulation, 1.94 for dexterity, 1.02 for emotion, 1.42 for cognition and 1.31 for pain. The EQ-5D-3L and EQ-5D-5L showed high values for mobility, self-care and usual activities and low values for anxiety/depression in comparison with other items. The HUI3 showed high values for ambulation and dexterity and low values for vision and hearing. Furthermore, the EQ-5D-5L showed a higher discrimination as

Table 2. Discrimination and difficulty

	Dimension	a	b1[a]	b2			
EQ-5D-3L	Mobility	4.58	-1.10	0.83			
	Self-care	4.33	-0.86	0.49			
	Usual activities	3.25	-0.76	1.32			
	Pain/discomfort	1.24	-2.51	0.26			
	Anxiety/depression	0.97	-3.41	0.29			
	Dimension	a	b1	b2	b3	b4	
EQ-5D-5L	Mobility	4.67	-0.79	-0.36	0.24	0.92	
	Self-care	6.69	-1.12	-0.57	0.02	0.72	
	Usual activities	3.39	-1.13	-0.28	0.52	1.42	
	Pain/discomfort	1.29	-3.42	-2.20	-0.87	0.74	
	Anxiety/depression	0.96	-4.83	-2.79	-1.16	1.01	
	Attribute	a	b1	b2	b3	b4	b5
HUI3	Vision	0.51	-11.30	-4.81	-4.01	-3.01	1.56
	Hearing	0.70	-4.48	-4.06	-2.87	-2.68	-2.48
	Speech	1.16	-2.87	-2.38	-0.40	-0.07	
	Ambulation	3.91	-1.14	-0.13	0.38	0.57	0.99
	Dexterity	1.94	-1.90	-1.10	-0.14	0.16	1.14
	Emotion	1.02	-4.77	-2.03	0.53	3.57	
	Cognition	1.42	-2.37	-0.59	0.57	1.19	1.40
	Pain	1.31	-3.52	-1.82	-0.53	0.96	

Abbreviations: a (discrimination), b (difficulty), EQ-5D-3L (EuroQol5 Dimension 3level), EQ-5D-5L (EuroQol5 Dimension 5level), HUI3 (Health Utilities Index Mark 3).

[a] EQ-5D-3L has three response levels and thus two difficulty-parameters, b1 and b2. b1 is the point of intersection of the level 2 item characteristic curve and level 3 item characteristic curve; b2 is the point of intersection of the level 1 item characteristic curve and level 2 item characteristic curve.

compared with the EQ-5D-3L and the HUI3.

Difficulty for the EQ-5D-3L was -1.10 and 0.83 for mobility, -0.86 and 0.49 for self-care, -0.76 and 1.32 for usual activities, -2.51 and 0.26 for pain/discomfort and -3.41 and 0.29 for anxiety/depression. The item characteristic curve of the EQ-5D-3L is shown in Figure 1. Difficulty of the pain/discomfort item was -2.51 (b1) and 0.26 (b2). The EQ-5D-3L has three response levels and thus two difficulty-parameters, b1 and b2 (b1 is the point of intersection of the level 2 item characteristic curve and level 3 item characteristic curve; b2 is the point of intersection of the level 1 item characteristic curve and level 2 item characteristic curve). Pain/discomfort was an easy item for stroke patients (negative θ represents low health states, and negative difficulty represents an easy category). Item characteristic curves for the EQ-5D-5L and HUI3 are not shown. Difficulty for the EQ-5D-5L was -0.79 to 0.92 for mobility, -1.12 to 0.72 for self-care, -1.13 to 1.42 for usual activities, -3.42 to 0.74 for pain/discomfort and -4.83 to 1.01 for anxiety/depression, and that for the HUI3 was -11.30 to 1.56 for vision, -4.48 to -2.48 for hearing, -2.87 to -0.07 for speech, -1.14 to 0.99 for ambulation, -1.90 to 1.14 for dexterity, -4.77 to 3.57 for emotion, -2.37 to 1.40 for cognition and -3.52 to 0.96 for pain. Difficulty was positively skewed only for usual activities for the EQ-5D-3L and for mobility and usual activities for the EQ-5D-5L. For the HUI3, difficulty was biased towards the negative for all items: in particular, vision, hearing and speech were heavily biased.

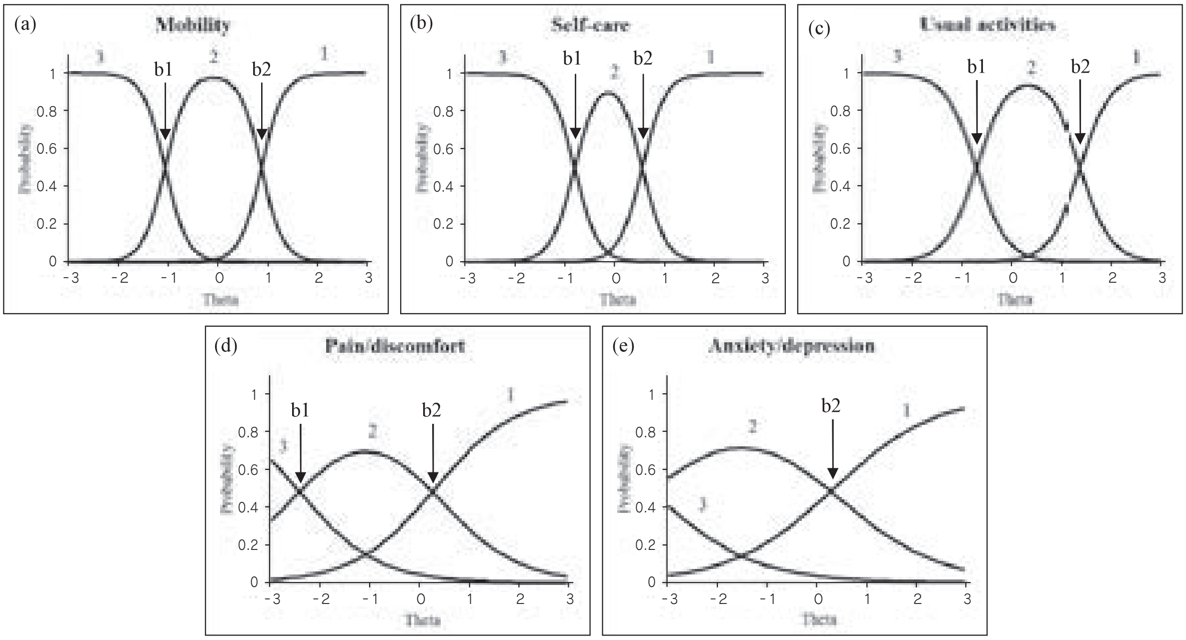

Figure 1. Item characteristic curves of the EQ-5D-3L
The vertical axis of the item characteristic curve shows the probability, and the slope of the curve indicates the item discrimination.The horizontal axis shows the difficulty (Ability of the subject), which indicates the mean health states (referred to as the population mean θ = 0). The EQ-5D-3L has three response levels and thus two difficulty-parameters, b1 and b2. b1 is the point of intersection of the level 2 item characteristic curve and level 3 item characteristic curve, and b2 is the point of intersection of the level 1 item characteristic curve and level 2 item characteristic curve.(a) Item characteristic curve for mobility, (b) item characteristic curve forself-care, (c) item characteristic curve forusual activities, (d) item characteristic curve forpain/ discomfort, and (e) item characteristic curve foranxiety/depression.Abbreviations: EQ-5D-3L (EuroQol5 Dimension 3level), b (Difficulty).

(2) Test characteristic curve and total information curve of the EQ-5D-3L, the EQ-5D-5L and the HUI3 (Figure 2 (a)–(c), Figure 3 (a)–(c))

The test characteristic curve of the EQ-5D-3L was a steep slope in the vicinity of θ = 1 and -1, but the slope was shallow for the other ranges. The curve of the EQ-5D-5L was also a steep slope in the vicinity from θ = 1 to -1. That of the HUI3 was shallow slope over a wide range, but it was a slightly steep slope in the vicinity from θ = 1.5 to -3.0.

The total information curve for the EQ-5D-3L was bimodal but showed a maximum of 13.1 at θ = -1. The curve for the EQ-5D-5L showed a maximum of 22.5 at θ = -0.6 and 10.0 or more at θ = -1.4 to 1.1. That for the HUI3 showed a maximum of 8.7 at θ = 0.5.

Discussion

We investigated the measurement properties of the HRQL scale for stroke patients using IRT with the EQ-5D-3L, EQ-5D-5L and HUI3 as the utility measurements.

1. Characteristics of the subjects and utility score

Diagnoses included 54.9% cerebral infarction cases, 35.0% cerebral hemorrhage cases and 7.6% subarachnoid hemorrhage cases. A similar breakdown was reported for Japan by the Ministry of Health, Labour and Welfare (about 60% cerebral infarction, 25% cerebral

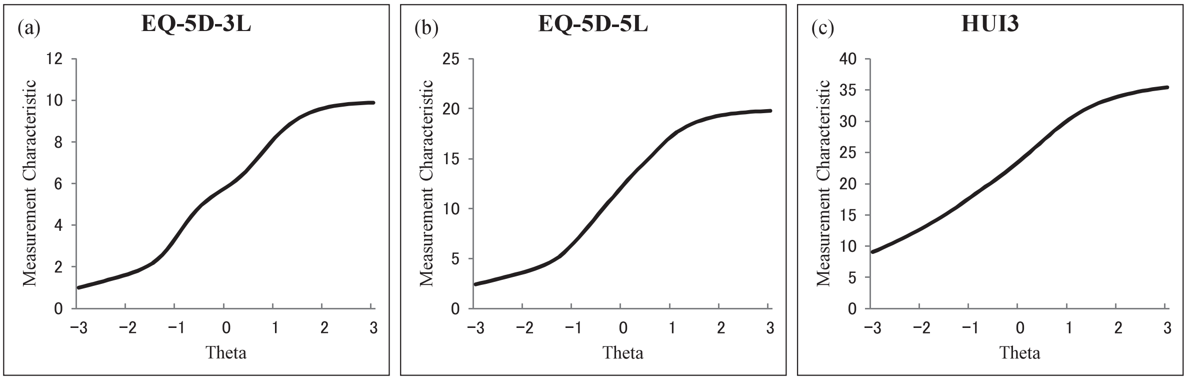

Figure 2. Test characteristic curves of the EQ-5D-3L, the EQ-5D-5L and the HUI3
The vertical axis of the test characteristic curve shows the total score (Characteristic), and the slope of the curve indicates the item discrimination.The horizontal axis shows the difficulty (ability of the subject), which indicates the mean health states (referred to as the population mean θ = 0).(a) Test characteristic curve for the EQ-5D-3L, (b) test characteristic curve for the EQ-5D-5L, and (c) test characteristic curve for the HUI3. Abbreviations: EQ-5D-3L (EuroQol5Dimension3level), EQ-5D-5L (EuroQol5 Dimension 5level), HUI3 (Health Utilities Index Mark 3).

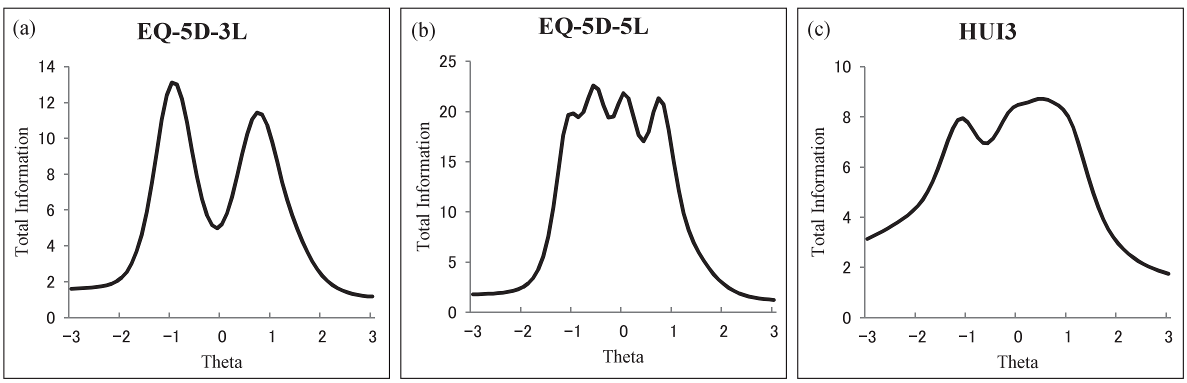

Figure 3. Total information curves of the EQ-5D-3L, the EQ-5D-5L and the HUI3
The total information curve represents the total the amount of information (Information) in the vertical axis and difficulty in the horizontal axis.(a) Total information curve for the EQ-5D-3L, (b) total information curve for the EQ-5D-5L, and (c) total information curve for the HUI3.Abbreviations: EQ-5D-3L (EuroQol5 Dimension 3level), EQ-5D-5L (EuroQol5 Dimension 5level), HUI3 (Health Utilities Index Mark 3).

hemorrhage and 10% subarachnoid hemorrhage cases) [23]. In terms of the MRS, Grade 3-5 patients accounted for about 70% of cases suggesting a relatively high degree of disability among the subjects. It is presumed that daily life without assistance is possible for Grades 0-2 and that subjects at Grade 3 or higher may be suited for rehabilitation assistance. The BI was 65.9 points suggesting a moderate degree of dependence [24]. This can be said to be a reasonable score for patients who cannot manage their own ADL in the recovery phase.

Utility scores for the EQ-5D-5L were lower than those for the EQ-5D-3L, and those for the HUI3 were the lowest among the three utility measurements. Single scores of the HUI3 showed low utility scores for ambulation (0.45), dexterity (0.64) and cognition (0.64). Noto et al [25] reported low utility scores for ambulation (0.32), dexterity (0.60) and cognition (0.54) in stroke patients. As such, characteristics of the subjects and utility scores were found to be close to the state generally found among stroke patients undergoing rehabilitation, which suggests that the disease characteristics were appropriate.

2. Discrimination and difficulty

The HUI3 showed low discrimination and difficulty for vision and hearing. This is thought to be due to the notion that damage to vision and hearing is unlikely to occur in the stroke patients who are the subjects of the present study. Discrimination was high except for vision and hearing, and we found that these can be determined with high accuracy even with a slight change in stroke patients. Discrimination for anxiety/depression was lower than other items. This is presumed to be because subjects were in the recovery phase, and because the period from onset was about 80 days, a stable mental state was likely to have been achieved.

For difficulty, the measurement accuracy in subjects with high health states was higher only for usual activities for the EQ-5D-3L, and for other items, we found that the measurement accuracy was higher for low subject health states. The EQ-5D-5L showed high measurement accuracy for usual activities and mobility in subjects with high health states. The cause of this difference was thought to be due to differences in the question sentence for mobility. In other words, whereas the text of the EQ-5D-3L, “I am confined to bed”, indicates a state whereby the subject imagines that no movement is possible, the text for the EQ-5D-5L is “I am unable to walk about”, and subjects can imagine that they are able to walk (move) a little bit. The curve of each item is biased toward the negative in all items in the HUI3, but discrimination was higher in the positive region. This is thought to be because utility scores used in the HUI3, which were measured for the general adult population, are biased [17]. However, the bias of discrimination and difficulty for the 3 measurements is not large, so there is no major effect when comparing the health states for stroke patients with the health states of healthy subjects and those with other conditions. Rather, it is shown that each measure, all of which are generic measurements, allows for an appropriate evaluation in stroke patients.

3. Test characteristic curve and total information curve of the EQ-5D-3L, EQ-5D-5L and HUI3

We found that the slope was steep for the test characteristic curve for the EQ-5D-3L (θ = 1 or -1), the EQ-5D-5L (θ = 1 to -1) and the HUI3 (θ = 1.5 to -3). Therefore, the EQ-5D-3L is suitable for subjects with a slightly higher or slightly lower health state, the EQ-5D-5L is suited for subjects with a slightly higher to slightly lower health state, and the HUI3 is suited for subjects with a slightly higher to lower health state. Furthermore, considering that the EQ-5D-5L has a steep slope in the range of 1 to -1 for θ, it is suited for subjects with health states in that range, and because the HUI3 has a moderate slope throughout, it is suited for other ranges.

On the basis of the total information curve, we found that the HUI3 and the EQ-5D-5L yield a stable amount of information. However, more information can be obtained in subjects with high health states because the θ for the HUI3 is positively skewed. For the EQ-5D-5L, more information is obtained in subjects with low health states because θ is skewed slightly negative.

Fryback et al [11] stated that measurement accuracy is high in subjects with a low health state for the EQ-5D-3L and the HUI3. Tokuda et al [10] reported that information peaked at a θ of -0.7 (information = 18.5) and decreased as θ increased. This difference is thought to be due to the fact that those studies covered a range of diseases and were not limited to stroke patients as in the present study. However, the test characteristic curve has a similar shape, so further validation is required. In the Tokuda et al study, as in the results of the present study, it was demonstrated that the measured characteristic differs depending on the measurement used. Therefore, the difference is thought to have occurred both as a result of the difference in the measurement used and the fact that the analysis was done in healthy subjects.

The finding that the EQ-5D-3L restricts the ability to discriminate small differences in health status has been questioned widely [26]. Moreover, a previous study reported a ceiling effect of the index component of the EQ-5D-3L [27]. Recently, a multi-country study generated evidence on the improved properties (reduced ceiling effect, improved discrimination) of the EQ-5D-5L compared with the EQ-5D-3L [28]. The present study supports the findings of this previous study in that measurement accuracy is higher because the amount of information from the EQ-5D-5L is higher than that from the EQ-5D-3L.

For future consideration, this study examined stroke patients, a group that is often subject to rehabilitation, but different findings are expected for different diseases. Therefore, research is also needed for orthopedic, tumor, respiratory and cardiovascular diseases. Moreover, because the health states are different depending on period of rehabilitation intervention considered, evaluation of different periods (acute phase, recovery phase and maintenance phase) is also needed.

Conclusions

This study demonstrated that utility measures including the EQ-5D-3L, EQ-5D-5L, and HUI3 can be used to measure a wide range of health states in stroke patients. However, because items with a low level of discrimination may be included depending on the measurement, it must be recognized that inclusion of such items might decrease the reliability of the overall measurement. Moreover, by simultaneously using the EQ-5D-5L and the HUI3 in stroke patients, accurate results can be expressed for a wide range of health states.

Acknowledgments

We would like to express our gratitude to all of the staff and patients of the hospitals where this research was conducted. This work was supported by MEXT KAKENHI Grant Number 25750195.

References

1. Fukuda T. Yakuzaikeizaigaku no kadai. Shakai Hoken Junpo. 2011; 2473: 13-18. (in Japanese)
2. Japanese EuroQol Translation team. The development of the Japanese EuroQol Instrument. J Health Care and Society. 1998; 8: 109-123. (in Japanese)
3. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L).Qual Life Res. 2011; 20: 1727-1736.
4. Ikeda S, Uemura T. Kenkoukouyouchisyakudo. In: Mandai T, editor. QOL hyouka manual. Tokyo: Intermedica; 2001: 56-65. (in Japanese)
5. Izumi R, Noto S, Uemura T, Sano T, Mizushima T, Sata T, et al. Reliability of Japanese version utility measures in health – related quality of life – agreement between self-reported and proxy-reported. Sogo Reha. 2011; 39: 569-575. (in Japanese)
6. Izumi R, Noto S, Uemura T, Sano T, Sata T. Validity and responsiveness of the health utility measures Japanese version in health-related quality of life: Evaluation of the use of EuroQol 5-Dimension and the Health Utilities Index Mark 3. Sagyo Ryoho. 2010; 29: 763-772. (in Japanese)
7. Bjorner JB, Petersen MA, Groenvold M, Aaronson N, Ahlner-Elmqvist M, Arraras JI, et al. Use of item response theory to develop a shortened version of the EORTC QLQ-C30

emotional functioning scale. Qual Life Res. 2004; 13: 1683-1697.
8. Orlando M, Sherbourne CD, Thissen D. Summed-score linking using item response theory: application to depression measurement. Psychol Assess. 2000; 12: 354-359.
9. Sasagawa S, Kanai Y, Muranaka Y, Suzuki S, Shimada H, Sakano Y. Development of a short fear of negative evaluation scale for Japanese using item response theory. Kodo Ryoho Kenkyu. 2004; 30: 87-98. (in Japanese)
10. Tokuda Y, Okubo T, Ohde S, Jacobs J, Takahashi O, Omata F, et al. Assessing items on the SF-8 Japanese version for health-related quality of life: A psychometric analysis based on the nominal categories model of item response theory. Value Health. 2009; 12: 568-573.
11. Fryback DG, Palta M, Cherepanov D, Bolt D, Kim JS. Comparison of five health-related quality of life indexes using item response theory analysis. Med Decis Making. 2010; 30: 5-15.
12. Ministry of Education, Culture, Sports, Science and Technology and Ministry of Health, Labour and Welfare. Available from: http://www.lifescience.mext.go.jp/files/pdf/37_139.pdf (accessed July 25, 2013) (in Japanese)
13. The EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy. 1990; 16: 199-208.
14. Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ. 2002; 11: 341-353.
15. EuroQol Group. Available from: http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Crosswalk_5L/EQ-5D-5L_Crosswalk_model_and__methodology.pdf (accessed July 22, 2013)
16. Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preference for health states. Oper Res. 1982; 30: 1043-1069.
17. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care. 2002; 40: 113-128.
18. Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. Md State Med J. 1965; 14: 61-65.
19. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. Stroke. 1988; 19: 604-607.
20. Kelly CW. Commitment to health scale. J Nurs Meas. 2005; 13: 219-229.
21. Fayers P, Hays R. Assessing quality of life in clinical trials. 2nd ed. New York: Oxford; 2005: 55-73.
22. Wim J, van der Linden WJ, Hambleton RK. Item response theory: brief history, common models, and extensions. In: van der Linden WJ, Hambleton RK, editors. Handbook of modern item response theory. New York: Springer; 1997: 1-28.
23. Ministry of Health, Labour and Welfare: Available from: http://www.mhlw.go.jp/topics/bukyoku/kenkou/seikatu/nousottyu/about.html (accessed July 25, 2013) (in Japanese)
24. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. J Clin Epidemiol. 1989; 42: 703-709.
25. Noto S, Uemura T. An improvement of a health utility score by rehabilitation and the validity of the Japanese version health utilities index. Iryo Keizai Kenkyu. 2009; 20: 73-84. (in Japanese)
26. Wu AW, Jacobson KL, Frick KD, Clark R,

Revicki DA, Freedberg KA, et al. 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.
27. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004; 13: 873-884.
28. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013; 22: 1717-1727.