

ORIGINAL REPORT

CONVERGENT VALIDITY AND RESPONSIVENESS OF THE EQ-5D UTILITY WEIGHTS FOR STROKE SURVIVORS

Wen-Shian Lu, PhD^{1,2}, Sheau-Ling Huang, MS^{3,4}, Jeng-Feng Yang, PhD⁵, Mei-Hsiang Chen, PhD^{1,2}, Ching-Lin Hsieh, PhD^{3,4} and Chia-Yeh Chou, MS^{3,6}

From the ¹School of Occupational Therapy and ²Occupational Therapy Room, Chung Shan Medical University Hospital, Taichung City, ³School of Occupational Therapy, College of Medicine, National Taiwan University, ⁴Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei City, ⁵Department of Physical Therapy, College of Medicine, National Cheng Kung University, Tainan City and ⁶Department of Occupational Therapy, College of Medicine, Fu-Jen Catholic University, New Taipei City, Taiwan

Objective: This prospective study examined the convergent validity and responsiveness of the EuroQoL-5 Dimensions Health Questionnaire (EQ-5D) utility weights in stroke survivors at the subacute stage.

Methods: Repeated assessments were conducted in rehabilitation wards of a medical centre. A study cohort was recruited from inpatients ($n=478$). A total of 342 participants (71.5%) completed 2 assessments; 1 at ward admission and 1 before hospital discharge. Outcome measures, including the Barthel Index (BI), Patient Health Questionnaire (PHQ-9) and EQ-5D, were administered at ward admission and before hospital discharge.

Results: The validity of the EQ-5D utility weights was assessed by comparison with the scores of the BI and PHQ-9. The absolute values of Pearson correlation coefficients ranged from 0.40 to 0.52 for the association of EQ-5D utility weights with BI scores and PHQ-9 scores at admission and before hospital discharge. The EQ-5D utility weights had a moderate effect size (0.76), moderate standardized response mean (0.74), and a significant difference (paired t -value=13.7, $p<0.001$) between baseline and follow-up in their ability to detect changes in the health status of stroke survivors staying in a rehabilitation ward.

Conclusion: This study supports the use of the EQ-5D utility weights as a valid and responsive instrument for performing cost-utility analyses of stroke survivors, including those with higher levels of dependence, at the subacute stage.

Key words: stroke; quality of life; psychometrics; utility weights; responsiveness; validity; EQ-5D.

J Rehabil Med 2016; 48: 346–351

Correspondence address: Chia-Yeh Chou, Department of Occupational Therapy, College of Medicine, Fu-Jen Catholic University, No. 510 Zhong Zheng Road, Xinzhuang Dist., New Taipei City, 24205 Taiwan. E-mail:068213@mail.fju.edu.tw

Accepted Jan 20, 2016; Epub ahead of print Mar 17, 2016

INTRODUCTION

Stroke is the second leading cause of death worldwide (World Health Organization (WHO), 2004) (1). Stroke survivors

usually have chronic disabilities and require long-term care and rehabilitation. Such healthcare demands and costs are currently of high concern, as the incidence of stroke increases with ageing populations. Thus, stroke-related cost analyses are important for determining priorities for healthcare policies and informing allocation of resources (2).

Economic evaluations can be performed using cost-utility analysis. Cost-utility analysis is frequently calculated in terms of the ratio of (total) costs and effects produced by the intervention or an alternative (3) and is increasingly estimated using Quality-Adjusted Life Years (QALYs). QALYs incorporate both survival years and impact on health-related quality of life (HRQoL) by calculating the utility weights in economic evaluations (2, 4, 5). Utility weights can be obtained by using the preference-based approach in terms of individuals' choices and preference ratings for various health states. A utility weight of 1 is representative of optimal health, and a weight of 0 is judged to be equivalent to death (4). Utility weights below 0 indicate health states perceived as being worse than death. Thus, utility weights can be used to represent people's subjective perspectives regarding the impact of their diseases or interventions. By adopting utility weights to calculate the QALYs, cost-utility analysis enables direct comparison of intervention outcomes and illness burden across diseases.

A generic HRQoL measure, such as the EuroQoL-5 Dimensions Health Questionnaire (EQ-5D), is necessary to achieve comparisons across diseases and therapies. National health services recommend using the EQ-5D (6). The EQ-5D has been translated for use in more than 170 countries/languages (6) and widely used across populations. The high frequency of its use may be attributed to the fact that the generic EQ-5D has the features of both generic utility measurements and a multi-dimensional health profile. Utility measurements are intended to summarize HRQoL levels into a single index value as utility weights. Moreover, the health profile is rated in various HRQoL dimensions, such as the physical, psychological, and social aspects. The EQ-5D can be used to monitor the heterogeneous characteristics of a multi-dimensional disease, such as stroke. Overall, using the EQ-5D in stroke survivors could provide economic evaluations to enable resource al-

location across diseases with the “same unit” (utility weight). That is, the ratings of the EQ-5D can be used to calculate the utility weights (EQ-5D utility weights) for further applications (e.g. comparisons of utility between different treatment protocols) (7).

The psychometric properties of the EQ-5D utility weights need to be validated for clinical and research use. The concurrent or convergent validity of a measure indicates the extent to which the measure assesses the construct that it is intended to assess (8). Sensitivity to change refers to the ability of an instrument to measure a change in state, regardless of whether the change is relevant or meaningful to a decision-maker (8). If a measure lacks evidence of validation of concurrent or convergent validity and sensitivity to change, users may not have confidence in selecting and using the measure. Because the validation of concurrent validity requires a gold standard measure as a criterion for comparison (9), a condition that cannot be met for the EQ-5D utility weights, the convergent validity and sensitivity to change of the EQ-5D utility weights should be validated.

Only one study with a sample of stroke survivors has validated the psychometric properties of the EQ-5D utility weights estimated with country-specific algorithms (10). In the study by Hunger et al., the psychometric properties of the EQ-5D were investigated in the first 6 months post-stroke (with a range of 4.1 to 10.1 weeks after stroke onset) and were re-assessed in a follow-up at 6 months. The results suggested that the concurrent validity and sensitivity to change of the EQ-5D were similar to those reported in the literature and validated in the patients at 1-year post-stroke onset. However, the psychometric properties of the EQ-5D utility weights are unclear for stroke survivors in the subacute stage, such as those with stroke onset of less than one month. Moreover, the stroke survivors enrolled in that study had at least a mild to moderate level of dependence in activities of daily living (ADL) (Barthel Index (BI) ≥ 35). The sample criteria limited the generalization of Hunger et al.’s findings. Thus, the results on the EQ-5D utility weights remain inconclusive, particularly considering the lack of recruitment of stroke survivors with a higher level of dependence. In summary, it is necessary to further investigate the validity and responsiveness of the EQ-5D utility weights in stroke survivors.

The aim of the current study was to validate the psychometric properties of the EQ-5D utility weights, specifically regarding their convergent validity and sensitivity to change, in stroke survivors in the subacute stage (within 1.5 months after onset) with various levels of dependence who were admitted to rehabilitation wards. Due to the lack of a gold standard measure, the convergent validity, rather than the concurrent validity, of the EQ-5D was to be validated. The results of this study provide empirical evidence on whether the EQ-5D utility weights are valid and responsive and thus useful in stroke survivors.

METHODS

Subjects

This prospective observational study was conducted from January 2009 to January 2012. Patients were recruited from a medical centre

in northern Taiwan (National Taiwan University Hospital, Taipei City) by consecutive sampling. Inclusion criteria were: (i) diagnosis of first intracerebral haemorrhage or ischaemic stroke confirmed either by computed tomography (CT) or by magnetic resonance imaging (MRI); (ii) ability to follow 1-step instructions; and (iii) currently in the subacute stage and receiving regular inpatient rehabilitation (i.e. 1 h occupational therapy, 1 h physical therapy, and 1 h speech therapy when necessary, per day, 5 times per week) at the medical centre. The inpatient rehabilitation programme was initiated when the patient exhibited a stable neurological condition and was transferred to the rehabilitation ward, usually 2–3 weeks after stroke onset.

Participants were excluded from follow-up if they: (i) had other central/peripheral neurological diseases, such as brain tumours or Parkinson’s disease, which could influence their motor control before or during recruitment; (ii) stayed in inpatient rehabilitation for less than 7 days (due to transfers or changes in condition); or (iii) did not give informed consent. Since 7 days would not be enough time for patients to receive sufficient rehabilitation or experience full natural recovery, a condition that which would lessen sensitivity to change, those with stays of less than 7 days were excluded to avoid bias in the results. The study protocol was approved by the Institutional Review Board of National Taiwan University Hospital, Taipei City. All participants provided written informed consent.

According to Frost et al. (11), a sample size of 200 is needed for psychometric studies. Thus, we planned to recruit more than 200 participants for this study.

Measures

The EQ-5D was developed by the EuroQoL group in 1990 (12). The EQ-5D is a generic and standardized measure of health and HRQoL status for clinical and economic appraisal. The EQ-5D comprises 5 subscales (dimensions), including mobility, self-care, usual daily activities, pain/discomfort, and anxiety/depression, with 3 levels of severity. When the EQ-5D is administered, the participants are asked to rate their HRQoL status in the 5 dimensions of the EQ-5D using 3 levels: “no problem (or 1),” “some/moderate problems (or 2),” and “severe problems (or 3)” (11, 12). This 3-level rating scale has the great advantage of being brief, allowing for quick administration. Moreover, the EQ-5D-3L index appears to be more responsive (Effect size [ES] 0.63–0.82; standardized response mean (SRM) 0.77–1.06) (13) than the EQ-5D with the revised 5-level rating scale (i.e. the EQ-5D-5L) (14) in stroke patients, while the convergent validity of the EQ-5D-5L is improved more than, or at least the same as, that of the EQ-5D-3L (15).

The EQ-5D utility weights were transformed from the ratings of the EQ-5D. Because the EQ-5D has 5 dimensions, each with 3 levels of severity, a total of 243 (3^5) possible health states can be defined, with 1, 1, 1, 1, 1 indicating perfect health status and 3, 3, 3, 3, 3 indicating the worst health status (16). The ratings can be calculated into a summary index score (i.e. utility weights). In this study, to calculate the utility weights, we adopted the Taiwanese calculation algorithms using the N3 model (level 3 occurred within at least 1 dimension) (7). For example, the ratings of the EQ-5D of 1, 1, 1, 1, 1, 2, 1, 1, 1, 3, 2, 2, 3, 3 and 3, 3, 3, 3, 3 were transformed into utility weights of 1.000, 0.766, –0.443 and –0.674, respectively (7).

The BI (17) is a commonly used measure of level of independence in ADL. The BI includes 10 items and has a total score ranging from 0 (worst) to 100 (best). The BI has been validated with good psychometrics for evaluating the ADL performance of stroke survivors (18, 19).

Since the domains of mobility, self-care, and usual daily activities of the EQ-5D can be categorized as “physical function” (which theoretically relates to ADL), the BI was selected as a reference instrument to further validate the convergent validity of the EQ-5D utility weights.

The Patient Health Questionnaire-9 (PHQ-9) (20) is a brief measure of depression and is derived from a self-administered version of the PHQ (21) of the Primary Care Evaluation of Mental Disorders (PRIME-MD) instrument for making a criteria-based diagnosis of depression. The PHQ-9 contains 9 depression-related items (and 1 item for rating severity of functional impairment). The PHQ-9 uses

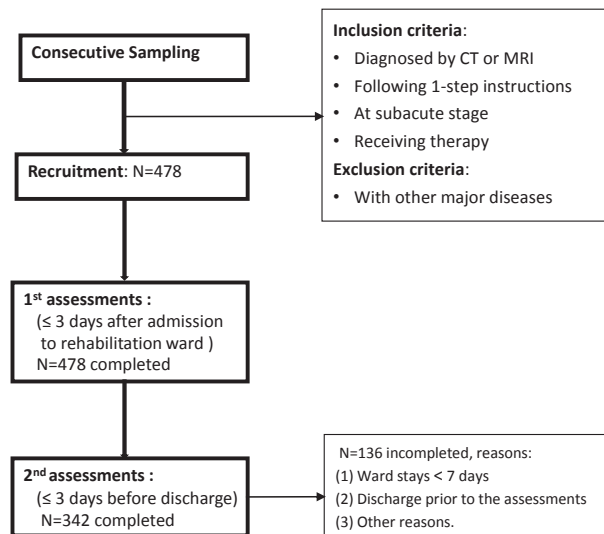


Fig. 1. Study procedure. CT: computed tomogram; MRI: magnetic resonance imaging; OT: occupational therapy; PT: physical therapy; ST: speech therapy.

the criteria for depression described in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV). Respondents are asked to rate the frequency of 9 depression symptoms that they have experienced within the past 2 weeks on a scale of “0” (not at all) to “3” (nearly every day) (16). The PHQ-9 has been validated as a screening and diagnostic instrument for assessing depression in stroke survivors (22).

Since the anxiety/depression domain of EQ-5D reflects the psychological aspect of health status and is theoretically related to “mood”, the PHQ-9 was selected as another reference measure to further validate the convergent validity of the EQ-5D utility weights.

Procedures

The assessments were administered at 2 time-points: within the first 3 days after admission to the rehabilitation ward, and within the 3 days before discharge. The EQ-5D and the PHQ were conducted by a research assistant using face-to-face interviews. The BI was administered by the patients’ physicians. In the study periods, the participants received routine physical therapy, occupational therapy, and speech therapy as needed. The therapies variously included training in postural control, ambulation, upper limb function, and ADL, and programmes for pain relief.

Statistical analysis

Convergent validity. Convergent validity was tested by correlating the EQ-5D utility weights with both BI and PHQ scores, using the Pearson correlation coefficient (*r*) and hypothesizing that we would find moderate correlations between variables. Convergent validity was considered large for Pearson’s $r \geq 0.60$, moderate for $0.31 \leq r \leq 0.59$, and small for $r \leq 0.3$ (9).

Sensitivity to change. Sensitivity to change refers to the ability of an instrument to measure a change in state, regardless of whether the change is relevant or meaningful to the decision-maker (8). The paired *t*-test statistic was calculated for the assessments administered at admission and within the 3 days before discharge. Also, the effect size (*d*) and standardized response mean (SRM) were used to analyse the sensitivity to change of the EQ-5D by comparing the change in the values of the EQ-5D utility weights obtained at admission and discharge. The *d* was calculated as $d = (S1 - S0) / SD_{baseline}$, where *S0* was the mean score of the baseline assessment, *S1* was the mean score of the follow-up assessment, and *SD*_{baseline} was the standard deviation (SD) of the baseline assessment. The method to calculate SRM was: $SRM = (S1 - S0) / SD$ of $(S1 - S0)$, where the numerator remained the same as for calculating *d*, and the denominator was the SD of the difference in the scores (23). According to Cohen’s interpretation guidelines, the absolute magnitude of the difference in the scores was considered to be trivial if the calculated *d* and SRM were < 0.20 , small if *d* and SRM ranged from 0.20 to 0.50, moderate if *d* and SRM ranged from 0.50 to 0.80, and large if *d* and SRM were ≥ 0.80 (24).

RESULTS

In total, 478 stroke patients were recruited and assessed at admission to the rehabilitation wards. Of these, 136 did not complete the follow-up assessments due to ward stays of less than 7 days (6 patients), discharge prior to assessment (73 patients), or other reasons. The flow chart of the study procedure is shown in Fig. 1. No significant differences in age, sex, BI score, PHQ-9 score, or EQ-5D utility weights were detected between patients who did and did not complete the assessments (*t*-test or χ^2 test, $p > 0.150$). A final total of 342 patients participated in the study, thus the requisite number of 200 participants for a psychometric study, as suggested by Frost et al., was exceeded (11).

Table I shows the sociodemographic and clinical characteristics of the patients who completed the evaluations at admission ($n = 478$) and follow-up ($n = 342$, 71.5%), and those who were

Table I. Characteristics of the stroke survivors

Characteristics	Admission (<i>n</i> = 478)	Lost to follow-up (<i>n</i> = 136)	Follow-up (<i>n</i> = 342)
Sex			
Male, <i>n</i> (%)	297 (62.1)	78 (57.4)	219 (64.0)
Age, years, mean (SD) [range]	65.3 (13.7) [19.7–97.1]	66.4 (14.5) [30.6–97.1]	64.9 (13.4) [19.7–95.3]
Time from onset to admission, days, median (IQR) [range]	16 (12–23) [4–213]	17 (12–24) [4–92]	16 (12–23) [4–213]
Length of stay, days, median (IQR) [range]	31 (20–44) [2–112]	28 (10–43) [2–112]	32 (22–44) [7–97]
EQ-5D utility weights, mean (SD) [range]	0.166 (0.359) [–0.674–1]	0.188 (0.392) [–0.674–1]	0.435 (0.356) [–0.674–1]
EQ-5D VAS, mean (SD) [range]	54.2 (24.5) [0–100]	54.8 (24.1) [0–100]	61.2 (21.3) [0–100]
BI score, mean (SD) [range]	33.9 (20.8) [0–90]	32.3 (25.0) [0–95]	71.1 (21.1) [0–100]
PHQ-9 score, mean (SD) [range]	14.9 (5.2) [7–31]	14.9 (4.8) [7–30]	13.3 (4.7) [8–30]

BI: Barthel Index; PHQ-9: Patient Health Questionnaire-9; VAS: visual analogue scale; IQR: interquartile range.

Table II. Convergent validity of the EuroQoL-5 Dimensions Health Questionnaire (EQ-5D) utility weights (Pearson correlation coefficients)

	EQ-5D utility weights	
	Admission ($n=478$)	Follow-up ($n=342$)
BI	0.42 ($p<0.001$)	0.52 ($p<0.001$)
PHQ-9	-0.40 ($p<0.001$)	-0.48 ($p<0.001$)

BI: Barthel Index; PHQ-9: Patient Health Questionnaire-9.

lost to follow-up ($n=136$, 28.5%). The mean scores of the BI were 33.9 at admission, 32.3 for lost to follow-up and 71.1 at follow-up. The mean scores of the EQ-5D utility weights/EQ-5D VAS were estimated as 0.166/54.2 at admission to the rehabilitation ward, 0.188/54.8 for lost to follow-up and 0.435/61.2 at follow-up (Table I).

Convergent validity. The values of the Pearson correlation coefficients between the EQ-5D utility weights and the BI scores at admission/follow-up were 0.42/0.52. The values of the Pearson correlation coefficients between the EQ-5D utility weights and the PHQ-9 scores at admission/follow-up were -0.40/-0.48 (Table II).

Sensitivity to change. The mean difference in the EQ-5D utility weights between the admission and follow-up was determined using the paired *t*-test and was significant (paired *t*-value= 13.7, $p<0.001$). The EQ-5D utility weights showed an effect size (*d*) of 0.76 and a SRM value of 0.74 between the 2 measurements over time.

DISCUSSION

The aim of this study was to validate the convergent validity and sensitivity to change of the EQ-5D utility weights in stroke patients. The observational cohort included a large sample of inpatients with stroke at the subacute stage. To date, the only other study to determine the convergent validity of the EQ-5D utility weights in stroke patients is that of ($n=210$) Hunger et al. (10). The participants in Hunger et al.'s study were more independent than those in our study, as indicated by their BI scores. In Hunger et al.'s sample, one-third of participants were assigned to phase C (BI scores 35–65) and two-thirds to phase D (BI scores above 65). In contrast, the mean BI score of our sample at baseline was 33.9. Considering this difference, our study makes a valuable contribution by expanding the generalizability of the convergent validity of the EQ-5D utility weights in stroke survivors to those who are less independent. In this study, the absolute values of the Pearson correlation coefficients (0.40–0.52) indicated moderate convergent validity of the EQ-5D. In comparison with Hunger et al.'s study (10) regarding the physical aspect of the EQ-5D, the correlations in the current study (0.42–0.52) are somewhat lower than the correlations of 0.55–0.67 between EQ-5D utility weights and physical summary score in the Stroke Impact Scale (SIS) (25) used by Hunger et al. (10). The primary reason for this discrepancy is that the constructs of the reference instrument

(the strength, hand function, mobility, and ADL subscale of the SIS) used to examine convergent validity by Hunger et al. are more similar to the construct (mobility, self-care and usual daily activities) of the EQ-5D than to the constructs of the reference instrument (BI) in our study. Regarding the psychological aspect of the EQ-5D, our results on convergent validity (0.40–0.48) are similar to those of Hunger et al. (0.32–0.44), implying that the main concepts and constructs of the 2 reference instruments (emotion subscale of SIS and Hospital Anxiety and Depression Scale (26)) in the former and current studies are alike. In summary, the EQ-5D utility weights are valid for stroke survivors at the subacute stage.

This study revealed that EQ-5D utility weights have a moderate degree of sensitivity to change, as demonstrated by the effect size (0.76) and SRM (0.74), while the SRM is independent of sample size. In contrast, in Hunger's study, sensitivity to change was tested using the effect size (*d*) only, and not SRM (10). The EQ-5D utility weights in that study had a small ES ($d=0.27$ –0.34). This discrepancy of degree of sensitivity to change may be addressed according to several issues.

The first issue concerns the sample characteristics, particularly regarding the potential for change. The illness stages of the stroke survivors in the 2 studies were different. Our stroke survivors were recruited at the subacute stage (i.e. mean/SD: 34 days/15.4 days after onset) and had the potential to show a greater change in health status than in those followed up at 6 months post-stroke (10). Also, the patients recruited in Hunger's study had BI scores of at least 35, and approximately two-thirds of the participants had BI scores greater than 65. As a result, that sample had limited potential for improvement, which could have led to underestimation of the sensitivity to change. In our study, the minimum BI score at admission was 13.1 (Table I), and the mean score of the BI at follow-up was 71.1. Consequently, the comparatively large improvement in independence in our participants resulted in a moderate level of sensitivity to change, higher than that found in a previous study (10). In other words, our results are more able to reflect the change in EQ-5D utility weights in stroke survivors at the subacute stage, including those with higher levels of dependence.

The utility weights of the patients improved. There are several possible explanations for this. First, patients with stroke undergo natural recovery at the subacute stage. In addition, the rehabilitation interventions at the subacute stage focused mostly on improving the patients' mobility, self-care, ADL, and on pain relief. The improvements in the aforementioned aspects may have decreased the patients' anxiety/depression. Thus, both natural recovery and rehabilitation may have contributed to the improvements in the utility weights in our study periods. Such a context may justify our study design for examining the sensitivity to change of the EQ-5D utility weights.

The second issue is the interval between 2 assessment points. In the study by Hunger et al., the researchers separated the study period into 2 intervals. The median time of the first interval was 10 days, a relatively short time between 2 assessments, which could lead to underestimation. Despite the second

interval of 6 months, the delayed pre-assessment at this interval could restrict the potential for improvement. This restriction could lead to underestimation of the sensitivity to change of EQ-5D utility weights. In the current study, the mean interval between 2 assessments was 34 days, and the time of investigation was more consistent with the period between admission and discharge. In this respect, our results more reliably reflect the sensitivity to change of EQ-5D utility weights in stroke survivors from admission to follow-up.

The third issue is related to the use of an external anchor for investigating sensitivity to change; specifically, Hunger et al. used the SIS as an external anchor. The methodology of using an external anchor to select the “changed patient” (e.g. the absolute score of the SIS subscale is greater than the smallest real difference in Hunger et al.’s study) may cause patients with a small improvement in EQ-5D utility weights to be overlooked, a shortcoming that could easily lead to overestimation of the results. Thus, our study, which did not employ an external anchor, seems to be acceptable because it used the entire cohort of participants to estimate sensitivity to change.

These factors may have resulted in a comparatively larger change in the utility weight scores in our participants. Consequently, the EQ-5D utility weights in the current study revealed a moderate level of sensitivity to change, higher than that in the previous study (10). In other words, the EQ-5D utility weights are able to reflect changes in the HRQoL in stroke survivors admitted to wards at the subacute stage.

Another issue is the adequacy of the sample size. In this study, we used statistical analyses, such as the Pearson correlation coefficient and SRM, to investigate the convergent validity and sensitivity to change of the EQ-5D utility weights. Since the results of these statistical methods are little influenced by sample size, to our knowledge, no conclusive method to choose the sample size of psychometric research has been suggested. Hopkins’ suggestion of 50 participants (27) and Frost’s suggestion of 200 cases (11) can be viewed as the lowest and the highest standards of a sufficient sample size. In this study, a total of 342 participants completed the follow-up assessments and were further analysed. The sample size was larger than both of the above standards; therefore the sample size of this study can be considered sufficient.

This study has several limitations. The first is the lack of a gold standard to test other psychometric properties, such as the concurrent validity of the EQ-5D utility weights, in stroke patients. Further studies are needed to validate the test-retest reliability of the EQ-5D utility weights for the estimation of random measurement error, which can help researchers and clinical practitioners determine whether the change in EQ-5D utility weights scores from admission to discharge is statistically significant or beyond random measurement error. In addition, we validated the sensitivity to change of the EQ-5D in this study. However, responsiveness, which represents the ability of an instrument to measure a meaningful or important change in a clinical status, is different from sensitivity to change (8). Further research to investigate the responsiveness of the EQ-5D

is warranted. Moreover, the findings of this study cannot be generalized internationally because the EQ-5D utility weights were estimated by using the country-specific algorithms established in Taiwan (7). Future studies of the utility of the EQ-5D would improve its utility in clinical routines and for cost utility analysis in the economic evaluation of stroke.

In conclusion, our findings show that the EQ-5D utility weights are a valid and responsive preference-based measure for patients with stroke staying in rehabilitation wards. The findings suggest that the EQ-5D utility weights can be used for performing cost-utility analyses of stroke survivors at the subacute stage.

ACKNOWLEDGEMENTS

This study was supported by National Health Research Institute (NHRI-EX102-10207PI).

The authors declare no conflicts of interest.

REFERENCES

1. World Health Organization. The global burden of disease: Geneva: WHO; 2015. 2004 update [cited 2015 Oct 10]. Available from: http://www.who.int/healthinfo/global_burden_disease/2004_report_update/en/.
2. Drummond M, Jonsson B, Rutten F. The role of economic evaluation in the pricing and reimbursement of medicines. *Health Policy* 1997; 40: 199–215.
3. van Exel NJ, Scholte op Reimer WJ, Koopmanschap MA. Assessment of post-stroke quality of life in cost-effectiveness studies: the usefulness of the Barthel Index and the EuroQoL-5D. *Qual Life Res* 2004; 13: 427–433.
4. Gold M, Siegel J, Russel L, Weinstein M. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.
5. Rasanen P, Roine E, Sintonen H, Semberg-Kontinen V, Ryyanen OP, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: a systematic literature review. *Int J Technol Assess Health Care* 2006; 22: 235–241.
6. EuroQol Group. EQ-5D: an instrument to describe and value health. Available from: <http://www.euroqol.org/>.
7. Lee HY, Hung MC, Hu FC, Chang YY, Hsieh CL, Wang JD. Estimating quality weights for EQ-5D (EuroQol-5 dimensions) health states with the time trade-off method in Taiwan. *J Formos Med Assoc* 2013; 112: 699–706.
8. Pardasane PK, Latham NK, Jette AM, Wagenaar RC, Ni P, Slavin MD, et al. Sensitivity to change and responsiveness of four balance measures for community-dwelling older adults. *Phys Ther* 2012; 92: 388–397.
9. Salter K, Jutai JW, Teasell R, Foley NC, Bitensky J, Bayley M. Issues for selection of outcome measures in stroke rehabilitation: ICF activity. *Disabil Rehabil* 2005; 27: 315–340.
10. Hunger M, Sabariego C, Stollenwerk B, Cieza A, Leidl R. Validity, reliability and responsiveness of the EQ-5D in German stroke patients undergoing rehabilitation. *Qual Life Res* 2012; 21: 1205–1216.
11. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD, Mayo FDAP-ROCMG. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 2007; 10: S94–S105.
12. EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16: 199–208.
13. Golicki D, Niewada M, Karlinska A, Buczek J, Kobayashi A,

- Janssen MF, et al. Comparing responsiveness of the EQ-5D-5L, EQ-5D-3L and EQ VAS in stroke patients. *Qual Life Res* 2015; 24: 1555–1563.
14. Herdman M, Gudex C, Lloyd A, Janssen M, 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.
 15. Golicki D, Niewada M, Buczek J, Karlinska A, Kobayashi A, Janssen MF, et al. Validity of EQ-5D-5L in stroke. *Qual Life Res* 2015; 24: 845–850.
 16. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; 35: 1095–1108.
 17. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J* 1965; 14: 61–65.
 18. Hsueh IP, Lee MM, Hsieh CL. Psychometric characteristics of the Barthel activities of daily living index in stroke patients. *J Formosan Med Assoc* 2001; 100: 526–532.
 19. Hsueh IP, Lin JH, Jeng JS, Hsieh CL. Comparison of the psychometric characteristics of the functional independence measure, 5 item Barthel index, and 10 item Barthel index in patients with stroke. *J Neurol Neurosurg Psychiatry* 2002; 73: 188–190.
 20. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; 16: 606–613.
 21. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA* 1999; 282: 1737–1744.
 22. Williams LS, Brizendine EJ, Plue L, Bakas T, Tu W, Hendrie H, et al. Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke* 2005; 36: 635–638.
 23. Konig HH, Born A, Gunther O, Matschinger H, Heinrich S, Riedel-Heller SG, et al. Validity and responsiveness of the EQ-5D in assessing and valuing health status in patients with anxiety disorders. *Health Qual Life Outcomes* 2010; 8: 47.
 24. Cohen J. *Statistical power analysis for behavioral science*. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
 25. Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The Stroke Impact Scale version 2.0. Evaluation of reliability, validity, and sensitivity to change. *Stroke* 1999; 30: 2131–2140.
 26. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983; 67: 361–370.
 27. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med* 2000; 30: 1–15.