Quality of Life Research (2019) 28:3177–3185
https://doi.org/10.1007/s11136-019-02254-1

Mapping the Shah-modified Barthel Index to the Health Utility Index Mark III by the Mean Rank Method

Yin Bun Cheung[1,2] · Hui Xing Tan[3] · Nan Luo[3] · Hwee Lin Wee[3,4] · Gerald C. H. Koh[3]

Accepted: 21 July 2019 / Published online: 27 July 2019

Abstract

Purpose To map the Shah-modified Barthel Index (SBI) to the Health Utility Index Mark III (HUI-3) in stroke patients, and to compare the performance of a recently developed method called the Mean Rank Method (MRM) against a popular method, the Ordinary Least Squares (OLS) method.

Methods A cohort of 473 patients who had their first clinical stroke diagnosis and hospital admission and were assessed using the SBI and HUI-3 at 3 months and/or 12 months post-admission. Observations were split to form a training dataset ($N=473$) and a validation dataset ($N=245$).

Results In the training dataset, the MRM using SBI total score as the predictor produced a mapped utility distribution that closely resembled the observed utility distribution. It had almost no shrinkage of the standard deviation ($P=0.542$), whereas the OLS using SBI total score and SBI item scores under-estimated the standard deviation by 28% and 26%, respectively (each $P<0.001$). The MRM mapping gave better fit in terms of smaller mean absolute error and larger intra-class correlation than the two versions of OLS mapping, whereas the OLS gave smaller mean-squared errors than the MRM. Multivariate regression analysis showed that the use of OLS-mapped utilities tended to under-estimate both the mean utility of people who had no comorbidity and the utility-comorbidity association as compared to the observed utility-comorbidity pattern although the differences did not reach statistical significance (each $P>0.05$). The MRM-mapped utility showed utility-comorbidity pattern more similar to the observed. Similar findings were obtained from the validation dataset.

Conclusions The MRM performed well. Mapping functions are available to map the SBI to the HUI-3 Utility Index.

Keywords Activities of daily living · Barthel Index · Health utility · Health Utility Index Mark III · Mapping · Stroke

Introduction

The utility of a health state can be measured by generic preference-based questionnaires such as the EuroQoL 5 Dimensions (EQ-5D) [1], Short Form-6D (SF-6D) [2], and Health Utilities Index Mark III (HUI-3) [3]. Cost-utility analysis, a common form of economic evaluation, is conducted to evaluate the value of adding a new health technology to the healthcare system. The value of the new healthcare technology is defined as the quality-adjusted life-years gained, i.e. survival gains weighted by the utility of the years lived [4].

Utility data is not always available. Mapping functions have been developed to map clinical measures such as health-related quality of life (HRQoL) and activities of daily living (ADL) to utility values [5, 6]. Essentially, mapping treats clinical measures as proxies of utility measures. It offers an opportunity to conduct cost–utility analysis when only clinical measures are available. Cost–utility analysis

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11136-019-02254-1) contains supplementary material, which is available to authorized users.

✉ Yin Bun Cheung
yinbun.cheung@duke-nus.edu.sg

1 Program in Health Services & Systems Research and Centre for Quantitative Medicine, Duke-NUS Medical School, Level 6, Academia, 20 College Road, Singapore 169856, Singapore

2 Centre for Child Health Research, Tampere University, Ylpönkatu 34, Tampere, Finland

3 Saw Swee Hock School of Public Health, National University of Singapore, 12 Science Drive 2, Singapore 117549, Singapore

4 Department of Pharmacy, National University of Singapore, 18 Science Drive 4, Singapore 117543, Singapore

based on mapping is accepted by healthcare authorities like National Institute for Health and Care Excellence (NICE), UK [7, 8], and Agency for Care Effectiveness, Singapore [9].

ADL assessment is often part of routine healthcare, health insurance policy and health services research. For example, in Singapore, claims under the ElderShield insurance scheme need to be substantiated by ADL assessment [10]. The Barthel Index (BI) is a widely used instrument for ADL assessment, including stroke patients [11]. It consists of 10 items, each measuring one aspect of ADL. It is recommended by the UK Royal College of Physicians for routine assessment of older people [12]. It had been mapped to the (3-level) EQ-5D Utility Index in stroke patients in the Netherlands [11] and in older people in the UK [12]. The Shah-modified BI (SBI) modified the 3-level classification of the BI to a 5-level classification. This improves the sensitivity and reliability of the measure [13, 14].

While all popular preference-based measures capture important aspects of health outcomes, the HUI-3 covers eight attributes, including speech and cognition [3]. These attributes may be more relevant to some diseases such as stroke and dementia. The HUI-3 may give a more comprehensive description of stroke patients. The BI or SBI have not been mapped to the HUI-3. They have conceptual overlapping such as in terms of ambulation and dexterity [13]. The conceptual overlapping provides a basis for mapping the SBI to the HUI-3.

Ordinary least square (OLS) is the most commonly used method for mapping clinical measures to health utility [15, 16]. One problem of OLS mapping is that it usually underestimates variance, and therefore, increases the risk of making type 1 errors in statistical hypothesis testing [5]. Another problem is that the application of OLS mapping functions tends to result in under-estimation of the utility of good health states and over-estimation of the utility of poor health states [17]. For example, a study of cancer patients showed that the mean OLS-mapped utility was lower than the mean of the observed EQ-5D utility values among patients with good clinical performance status, whereas the opposite error was found in patients with poor performance status [18]. In a study of people living with dementia, mean OLS-mapped utility was lower than the observed mean HUI-3 utility in patients with good clinical status as per the Clinical Dementia Rating scale, and vice versa [19].

Other regression-based methods have been investigated as potential alternatives to OLS. Some empirical comparisons showed OLS performed better than all these alternatives while others showed some of these alternatives performed better than OLS and each other. For example, in a study of cancer patients that our research group conducted to map the Functional Assessment of Cancer Therapy-General (FACT-G) to the EQ-5D utility, we found that the censored least absolute deviation (CLAD) regression was more accurate than the OLS [18]. However, in a study of breast cancer patients also conducted by our group, the OLS performed well as compared to CLAD, Tobit, quantile regression and logistic quantile regression [20]. We concluded that: "Although its theoretical assumptions may not be valid, ordinary least square outperformed other regression methods [20]." Similar conclusion was made by our mapping study in Parkinson's disease [21]. In a community-based study, Sullivan and Ghushchyan found that the OLS and CLAD gave very similar mapping results even though 46% of the utility values were at the ceiling level, a condition that theoretically should be favorable to CLAD and unfavorable to OLS [22]. The MM-estimator was found to perform well in a pediatric study [23] but not in a study of adults with urinary incontinence [24]. In a cancer study, multinomial logistic regression was found to give the best fit for mapping the EORTC QOL-C30 to EQ-5D, whereas OLS was the second best [25]. But the same study showed that OLS was the best for mapping the FACT-G to EQ-5D utility, whereas the multinomial logistic regression and two-part models gave the poorest results [25]. A study of HIV patients that aimed to map the Medical Outcomes Study-HIV Health Survey to EQ-5D and HUI-3 utility values found that a finite mixture model gave the best fit to EQ-5D but OLS gave the best fit to HUI-3 [26]. These two studies are exemplary: The various regression-based challengers collectively won in half of the applications; OLS single-handedly won in the other half [25, 26]. Taking these evidences together, the OLS appears accurate and robust for utility mapping as compared to other regression-based alternatives. Regression-based methods do not appear to be a promising direction to improve mapping accuracy.

A new mapping procedure, called the Mean Rank Method (MRM), was recently proposed [17]. Its performance has been demonstrated in both simulations and applications to map the Functional Assessment of Cancer Therapy-Breast and the World Health Organization Quality of Life-BREF (WHO-QOL-BREF) to the (5-level) EQ-5D Utility Index [17, 27] and the Alzheimer's Diseases Cooperative Study—Activities of Daily Living Inventory to the HUI-3 [19]. It tended to perform better than the equipercentile method [5] and, depending on what evaluation criteria were used, was superior or non-inferior to the OLS method.

The present study aimed to map the SBI to the HUI-3 utility index among stroke patients in Singapore. We also demonstrated the use of the MRM and compared it with the OLS method.

Methods

Participants and measurements

Inpatients aged over 40 years who had their first clinical stroke diagnosis and were not globally aphasic were recruited from 2010 to 2013 from five main public tertiary hospitals of Singapore, including Changi General Hospital, Khoo Teck Puat Hospital, Tan Tock Seng Hospital, Singapore General Hospital, and National University Hospital. The study was originally planned to assess the health, social and financial effects of caregiving among caregivers of stroke patients in the 1st-year post-stroke, with follow-up assessments at 3 and 12 months post-admission. Among eligible participants, the recruitment rate was 38.3%. Comparison of baseline characteristics of participants present and missing at the two follow-up assessments showed that marital status was the only variable that had a statistically significant difference between the groups ($P < 0.05$ at both timepoints). Participants who were married were less likely to be lost to follow-up. The study was approved by the National University of Singapore Institutional Review Board, SingHealth Centralized Institutional Review Board and the National Health Group Domain Specific Review Board. Written informed consent was obtained from both the patients and the caregivers after explaining the purpose of conducting the study as well as the procedures involved. The present secondary analysis of anonymized data was approved by the National University of Singapore Institutional Review Board (S17-257E).

The SBI, HUI-3 and other clinical and health variables such as Charlson's Comorbidity Index (CCMI) were collected at baseline during index stroke admission and 3 months and 12 months post-admission. However, the HUI-3 and SBI collected at baseline referred to different timepoints that represented pre- and post-stroke diagnosis. Therefore, in this study, only data at the 3- and 12-months post-admission assessment was used. Out of 661 cohort members, 412 and 308 participated in the 3- and 12-month assessments, respectively.

HUI-3 is a generic preference-based system for measuring health outcomes and producing utility values [3]. It measures 8 attributes, namely, vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. A multiplicative multi-attribute utility function is available to convert the responses to the 8 descriptors into a health utility [28]. It has a possible range of −0.359 to 1. The HUI-3 has been validated in the previous studies [29, 30].

Statistical analysis

The MRM is conceptually similar to the equipercentile method. It uses a simple procedure to handle tied values without requiring smooth cumulative distribution functions [17]. It uses only one predictor. In the present study, the MRM maps the SBI total score (sum of item scores) to the HUI-3 utility. The procedure is as follows:

1. The SBI total scores are sorted in ascending order. A rank is assigned to each observation. For observations with tied SBI total scores, the mean of ranks is assigned.
2. The HUI-3 utility values are sorted in ascending order. A rank is assigned to each observation. The ranking among observations with tied HUI-3 utility values is arbitrary.
3. Each unique SBI total score is mapped to the HUI-3 utility value that has the same rank.
4. In the case of n tied SBI total scores, they are mapped to the mean of the n consecutive HUI-3 utility values whose mean of ranks equals the mean of ranks of the tied BI scores.

The mapped utility is referred to as MRM-total in this article.

For the implementation of OLS mapping, we used the fractional polynomial (FP) approach with one or two power terms (FP1 model or FP2 model) to capture the possibly non-linear relation between SBI total score and HUI-3 [31]. We also used the FP approach to simultaneously map multiple SBI items to HUI-3, with stepwise backward selection method and $P < 0.1$ as the selection criterion. The OLS-mapped values are referred to as SBI total and SBI items, respectively. We also included linear OLS models for comparison. The R-squared and deviance difference as well as graphical examinations were used to guide model selection.

We did not include covariate adjustment in the regression analysis since this is a not a study of causation. There is no confounder to consider in the usual epidemiology sense. We are aware that some studies included covariates in the mapping functions. We do not agree with this practice because (a) it changes the purpose of mapping (to map a descriptive score to utility, not to map multiple variables to utility), and (b) it reduces the practical value of the mapping functions developed (they become not usable unless the covariate values are also available).

Training and validation datasets

Data were collected at the 3- and 12-month timepoints. For participants who had only one measurement, the data was used for the training/development of mapping functions. For participants who had two measurements, we randomly selected one measurement for training and used the other for validation. There are other methods of splitting independent observations into training and validation datasets. However, they are not developed for correlated data, such as in the present study where one participant may contribute more

than one observation. The way we split the data prevents statistical difficulty in the analysis of correlated data.

Evaluation criteria

The choice of evaluation criteria can affect the conclusion. The mean-squared errors by definition favors OLS mapping. To obtain a balanced overview, we used multiple criteria to assess the usefulness of each method. First, we compared the distribution of the observed and mapped utilities, including percentiles, mean and standard deviation. Second, we calculated measures of individual level agreement between the mapped and observed utility values, including mean-squared errors (MSE), mean absolute errors (MAE), intra-class correlation (ICC) and the square of Pearson's correlation coefficient (R-squared, or R^2). Third, we examined whether the mapped utilities could be used to reproduce the association estimates between the observed utilities and Charlson's Comorbidity Index (CCMI), by multivariate regression analysis, which allows the testing of equal regression coefficients between regression equations with observed and mapped utilities as dependent variables [32].

Results

Participant characteristics

The number of participants who were assessed for both SBI and HUI-3 at 3-month, 12-month and both timepoints were 165, 63 and 245, respectively, giving a total of 718 observations. The training and validation dataset included 473 and 245 observations, respectively.

The characteristics of the participants in the training and validation datasets are shown in Table 1. In the training dataset, the participants were on average 63-year-old and about two-thirds of them were male. The ethnic composition approximately represented the stroke patient population in Singapore, with 65%, 26% and 8% ethnic Chinese, Malay and Indian, respectively [33]. Most strokes were infarctions; comorbidity was common (mean CCMI = 5). The mean HUI-3 and BI were 0.59 and 84.9, respectively. The percentages reaching the HUI-3 and BI ceilings were 5.5% and 53.5%, respectively. The validation dataset had similar participant profile.

Table 1 Participant characteristics

Characteristic	Training (N = 473)		Validation (N = 245)	
	N/mean	%/SD	N/mean	%/SD
Age, mean (SD) (years)	62.7	10.9	61.7	10.4
Gender				
Male	316	66.8	160	65.3
Female	157	32.4	85	34.7
Ethnicity				
Chinese	306	64.7	150	61.2
Malay	123	26.0	68	27.8
Indian	37	7.8	24	9.8
Others	7	1.5	3	1.2
Type of stroke				
Infarction	414	87.5	213	86.9
Hemorrhage	43	9.1	24	9.8
Both	12	2.5	6	2.5
Missing	4	0.8	2	0.8
CCMI, mean (SD)	5.0	1.7	5.1	1.6
HUI-3, mean (SD)	0.59	0.36	0.62	0.33
At HUI-3 ceiling	26	5.5	16	6.5
At HUI-3 floor	0	0.0	0	0.0
SBI, mean (SD)	84.9	27.5	86.6	25.2
At SBI ceiling	252	53.3	136	55.5
At SBI floor	4	0.9	2	0.8

Number (N) and % for categorical variables; mean and standard deviation (SD) for continuous variables

Development of mapping functions

We began with using the SBI total score as predictor and implementing the OLS by 2-degree FP (FP2). Deviance difference showed that it fitted better than 1-degree FP ($P < 0.001$) and a linear OLS model ($P < 0.001$). The FP2 model gave R^2 of 0.535. However, this model showed an implausible, non-monotonic pattern, where the predicted SBI values in the SBI range of about 30 to 60 were slightly lower than the poorer BI values in the range of about 0 to 30 (Fig. 1).

In contrast, despite statistical significance, the 1-degree FP (FP1) model only gave a slightly smaller R^2, at 0.514, than the FP2 model. Importantly, it showed a monotonic pattern where lower SBI scores predicted lower HUI-3 utility values (Fig. 1). It also fitted better than a linear OLS model ($P < 0.001$), which gave an R^2 of 0.429. Therefore, we chose the FP1 model as the best OLS mapping model and, unless specified otherwise, OLS-total refers to the utility value generated by OLS FP1 model with the SBI total score as predictor. The estimated OLS FP1 mapping equation is:

$$\text{HUI-3 utility} = 0.043638 + 6.97 \times 10^{-7} \times (\text{SBI} + 1)^3$$

An OLS model with SBI items as predictors was also implemented. The stepwise backward selection procedure retained 5 items, including dressing, feeding, toileting, ambulation/wheelchair and stairs. The model gave an R^2 of 0.548. The utility thus generated is referred to as OLS-items in this article. The Model details are provided in Spreadsheet 1 of Online Supplementary File 1. In contrast, an item-level

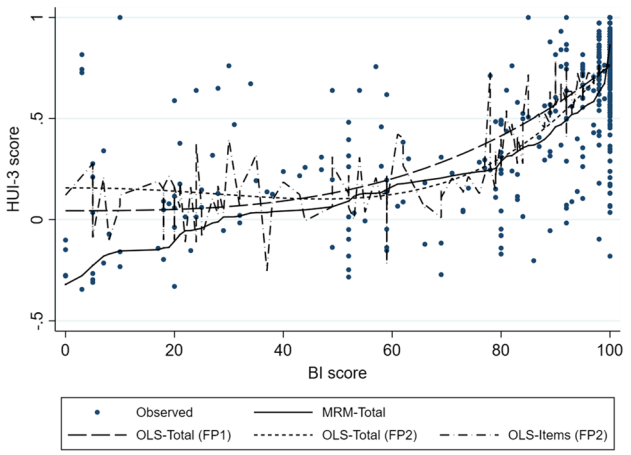

Fig. 1 Observed and mapped HUI-3 utilities in relation to Shah-modified Barthel Index. MRM-total: mapping of BI Total score to HUI-3 by MRM; OLS-Total/Items: mapping of BI total/item scores to HUI-3 by OLS. FP1 and FP2: 1- and (up to) 2-degree fractional polynomials

OLS model that assumed linear relation and used the same stepwise backward selection procedure gave an R^2 of 0.481. When the OLS-items values were plotted against the SBI total score, they mainly fluctuated around the OLS FP1 model, with the exception that they were closer to the OLS FP2 model in the range of 65 to 75 (Fig. 1).

The MRM mapping function with SBI total score as predictor is also shown in Fig. 1 (MRM-total). It increased monotonically with higher SBI scores. The OLS-total, OLS-items and MRM-total gave similar mapped utility values near the middle and top SBI values (about 60 and 100, respectively). At the lower range of SBI values, the OLS-mapped values tended to be higher than the MRM-mapped values. The MRM-total mapping function is provided in Spreadsheet 2 of Online Supplementary File 1.

Evaluation in training dataset

The upper panel of Table 2 shows the distribution of the observed and mapped utilities in the training dataset. The means of all mapped utility values were identical to the observed mean, 0.590. The standard deviation (SD) of the MRM-total was lower than the observed by 2%. OLS-total and OLS-items under-estimated the SD by about 28% and 26%, respectively (Pitman's test, each $P < 0.001$). The MRM gave minimum, 10th, 25th, 75th and 90th percentiles and maximum closer to the observed values than the OLS did.

Table 3 shows the measures of agreement between the observed and mapped utilities at the individual level. OLS-items had the smallest mean-squared errors and largest R^2, but MRM-total had the smallest mean absolute errors and largest ICC.

To investigate whether the mapped utilities could be used to reproduce the association estimates between the observed utilities and CCMI, we examined the gradient of HUI-3 utility values in relation to comorbidity. In the observed HUI-3 data, a linear trend across CCMI fitted well as compared to CCMI as a categorical variable (likelihood ratio test, $P = 0.765$). Table 4 compares the regression findings using the multivariate regression method.

Table 3 Mean-squared errors (MSE), mean absolute errors (MAE), intra-class correlation (ICC) and R-squared (R^2) of mapped HUI-3 utilities compared to observed utilities in the training and validation datasets

Dataset	Utilities	MSE	MAE	ICC	R^2
Training	MRM-total	0.068	0.183	0.727	0.528
	OLS-total	0.062	0.188	0.679	0.514
	OLS-items	0.057	0.185	0.708	0.547
Validation	MRM-total	0.061	0.171	0.733	0.537
	OLS-total	0.056	0.179	0.672	0.502
	OLS-items	0.056	0.178	0.674	0.504

MRM-total mapping SBI total score to HUI-3 by MRM, *OLS-total/items* mapping SBI total/item scores to HUI-3 by OLS

Table 2 Distribution of observed and mapped HUI-3 utilities in training and validation datasets

Dataset	Utilities	Mean	SD	Min	P10	P25	P50	P75	P90	Max
Training	Observed	0.590	0.356	−0.344	0.034	0.333	0.712	0.866	0.973	1.000
	MRM-total	0.590	0.350	−0.320	0.034	0.336	0.862	0.862	0.862	0.862
	OLS-total	0.590	0.255	0.044	0.076	0.457	0.762	0.762	0.762	0.762
	OLS-items	0.590	0.264	−0.252	0.128	0.417	0.765	0.765	0.765	0.789
Validation	Observed	0.625	0.335	−0.333	0.082	0.429	0.733	0.892	0.973	1.000
	MRM-total	0.608	0.338	−0.320	0.091	0.369	0.862	0.862	0.862	0.862
	OLS-total	0.604	0.242	0.044	0.147	0.503	0.762	0.762	0.762	0.762
	OLS-items	0.610	0.242	−0.195	0.186	0.492	0.765	0.765	0.765	0.976

P percentile, *MRM-total* mapping SBI total score to HUI-3 by MRM, *OLS-total/items* mapping SBI total/item scores to HUI-3 by OLS

While there was no statistically significant difference, OLS-total most under-estimated the intercept and slope, followed by OLS-items. The MRM most closely reproduced the association pattern.

Evaluation in validation dataset

The lower panels of Tables 2, 3 and 4 show the evaluation results on the MRM- and OLS-mapped utilities in the validation dataset. The findings were largely similar to those obtained from the training dataset, except that MRM now more clearly out-performed OLS.

In the lower panel of Table 2, the OLS-items predicted a maximum of 0.976, very close to the observed maximum 1.0. However, there was only one observation predicted to have this value. The second higher prediction was 0.765, which was the same as the predicted 50th, 75th and 90th percentiles. The single observation with mapped utility 0.976 had an atypical pattern of limitations on ADLs, with the most serious limitation in toileting but no limitation in the other ADLs selected for inclusion in the OLS-items mapping. As can be seen in the OLS-items mapping function, there were non-linear relations between the SBI scores on feeding and toileting and HUI-3 utility. This combination of an atypical ADL pattern and non-linearity created this outlier. The true HUI-3 utility for this observation was 0.707. In the validation dataset, the differences in MSE, MAE, ICC and R^2 between OLS-items and OLS-total clearly narrowed (Table 3).

In the validation dataset, MRM-total had the largest R^2 in addition to the largest ICC and smallest MAE (Table 3). Furthermore, its association with CCMI more closely reflected the observed pattern while the OLS utility values more clearly under-estimated the intercept and slope (Table 4).

Discussion

This study extends our knowledge in two ways. First, it has provided MRM and OLS mapping functions for conversion of SBI values to HUI-3 utility values. This can facilitate health technology assessment where national agencies accept utility data generated by mapping. Second, it has strengthened the emerging evidence about the performance of the MRM.

Two previous studies had mapped the (original) BI to the (3-level) EQ-5D. Since both the predictor and the utility measure are different from this study's, comparison of findings is only approximate. In the UK study of older people, the MAE of different mapping models were about 0.20 to 0.25 in various models [12]; in the Dutch study of stroke patients, the ICC was 0.70 [11]. The MRM in the present study gave similar level of accuracy, with MAE about 0.18 and 0.17 in the training and validation datasets, respectively, and ICC about 0.73 in both the training and validation datasets. The UK and Dutch studies only map to a maximum utility of 0.79 and 0.75, respectively [11, 12]. The proposed MRM function has less ceiling effect (0.862).

In this study, about half of the observations reached the SBI ceiling at 100. Therefore, the MRM- and OLS-mapped utility values above the 50th percentile did not vary (except that single OLS-items outlier). Nevertheless, at the individual level, a higher SBI score is mapped to a higher HUI-3 utility according to the MRM function, as seen in Fig. 1 and Online Supplementary File 1. This is true even for a change from SBI score 99 to 100 (or from 1 to 0). Thus, the function can be used to measure improvement (or deterioration) as long as SBI is not 100 (or 0) to begin with. It is foreseeable that in some context many study participants may have SBI score 100 from the beginning. In such a situation, the use of SBI-based

Table 4 Regression analysis of observed and MRM- and OLS-mapped HUI-3 utilities in relation to Charlson's Comorbidity Index (CCMI) in training and validation datasets

Dataset	Covariate	Utilities	Intercept	P^a	Slope	P^a	Model P^b
Training	CCMI	Observed	0.708		−0.024		
		MRM-total	0.673	0.345	−0.017	0.317	0.606
		OLS-total	0.657	0.144	−0.013	0.123	0.303
		OLS-items	0.663	0.180	−0.015	0.156	0.366
Validation	CCMI	Observed	0.870		−0.048		
		MRM-total	0.872	0.963	−0.052	0.699	0.529
		OLS-total	0.791	0.114	−0.037	0.226	0.182
		OLS-items	0.779	0.068	−0.033	0.107	0.173

MRM-total mapping SBI total score to HUI-3 by MRM; *OLS-total/items* mapping SBI total/item scores to HUI-3 by OLS

[a]Test of difference in intercept or slope between mapped versus observed data

[b]Joint test of differences in intercept and slope

mapping to evaluate a health promoting intervention may lead to over-estimation of cost–utility ratio. In this situation, it is advisable to collect utility data instead of relying on mapping, as HUI-3 has less ceiling effect than the SBI.

The application of OLS mapping functions suffers several known problems [5, 17]. Other regression-based mapping methods occasionally performed worse and occasionally performed better than the OLS [18, 20–26]. There is no consistent result about their relative performance. Given the lack of progress based on regression methods, the equipercentile was proposed as a promising alternative to regression-based mapping [5]. But it was recognized that the equipercentile was difficult to implement in health studies due to discrete distributions and ceiling effects. In both simulations and in real datasets [17, 27], both the MRM and OLS out-performed the equipercentile method.

OLS and other regression-based methods have an advantage that it is straight-forward to include multiple predictors and therefore suitable for mapping items of ADL or HRQoL measures to utility measures. Item-level mapping that assumed linear relation is well-documented in the literature. The issue of non-linear relation in item-level mapping has been less discussed. In our OLS mapping of the SBI total score to the HUI-3, we preferred a 1-degree FP model over a 2-degree FP despite slightly lower R^2 in the former, because it showed a more plausible monotonic pattern of utility increased with higher SBI. In the item-level OLS mapping, we lost the ability to maintain this preference. This is because if we choose a pattern for one item, the other items may change from a monotonic to non-monotonic pattern. Indeed, two items, feeding and toileting, showed a non-monotonic relation with HUI-3 utility. This contributed to the outlier, a high mapped utility value (0.976) for a person who had serious limitation in toileting in the validation data. The OLS-items appeared to be more capturing idiosyncratic data features in the training dataset than OLS-total and MRM. On the other hand, using item-level analysis with only linear relation gave worse model fit (smaller R^2). The present work has highlighted a feature of the MRM: The mapping function is constrained to be monotonic. This is often a conceptual advantage as reversing trajectories may not be plausible.

A limitation of the MRM is that it only uses one predictor, the SBI total score in this case. Nevertheless, in empirical comparison with OLS that used either a total score or item scores, it still performed well. One possible extension of the MRM is that, when there are two predictors, use one predictor to rank the observations and then use the other predictor to rank among the tied observations. For example, some HRQoL measures generate a physical summary score and a mental summary score. Since utility measures often include more physical/functional than mental dimensions, investigators may consider using the physical summary score in the first level of ranking and mental summary score in the second level. Another limitation is that we have focused on comparing MRM and OLS and have not included comparisons between regression-based methods.

One more limitation of the present study is that the validation only used internal data. The training and validation datasets may be relatively similar as they come from the same study, which could mean relatively favorable validation results. This could have contributed to the performance of the MRM and OLS mapping functions in the validation dataset. However, it should not affect the relative performance of MRM and OLS methods. Further validation of the mapping functions using external data will be valuable.

Whether the MRM out-performed the OLS depended on evaluation criteria used. At the group level, the MRM performed better than the OLS in terms of preserving the spread of utility distribution and preserving association pattern with clinical covariates [17, 19]. In terms of mapping errors at the individual level, the MSE by definition favors the OLS. When other indicators are used, such as MAE and ICC, there has been no consistent pattern. In this study the MRM has better accuracy at the individual level according to most of these indicators, but previous studies have shown variable results [17, 19, 27]. We speculate that the range of utility values is a contributor to the relative performance of different methods. A previous study suggested that OLS mapping of utility in a limited range may be more accurate than in a broad range [34]. The previous breast cancer study that showed better accuracy of OLS as compared to MRM had no negative utility values and most of the observed utility values were above 0.5. The present and two other previous studies that showed less accuracy of OLS as compared to MRM had broader ranges and covered negative utility values.

Our study involved patients within 12 months of new stroke diagnosis in Singapore. The data covered the whole range of SBI values. Users of the mapping results can apply it without restriction in this regard. The cohort study did not record reasons for loss to follow-up. So we could not provide a complete description of panel attrition and its potential impact. Nevertheless, the study recruited participants from five major hospitals in different parts of Singapore. The multi-centre nature and recruitment criterion ensured that the sample findings are generalizable to patients who recently have had stroke in the country. Furthermore, comparing this sample with patients attending a rehabilitation clinic in Singapore who were on average 22 months post-stroke [35], it can be seen that their age, gender and SBI are not very different: mean SBI equaled 88.9 in the rehabilitation study versus 84.9 and 86.6 in the training and validation samples here. Further comparing with data from the National Disease Registry Office [33], which pooled first and recurrent strokes, our sample

bore some characteristics that are similar to the national data, such as older age among Chinese than Malay and Indian patients, and older age among females than males. As such, we expect that the findings from this sample are widely applicable to stroke patients in Singapore in general. Further validation of the mapping functions in other populations would help to shed more light on the performance of the mapping functions generated and generalizability of the comparative findings.

Currently, there is no Singapore tariff for the calculation of the HUI-3 utility. Hence, we used the original, Canadian tariff. The HUI-3 system together with the Canadian tariff had previously been shown to be valid in healthcare settings in Singapore [29, 30]. The Agency for Care Effectiveness has indicated that local tariff is preferred if available, but it is not a requirement [9]. Therefore, the mapping functions developed here will be useful in Singapore. Furthermore, it may also be useful in other countries where there is no local HUI-3 tariff, but further evaluation will be needed to confirm this.

There is a number of strengths in this study. The mapping functions generated are applicable to the full range of the SBI scores. As compared to many other studies that are single-centre, this multi-centre study has the strength of wider generalizability. Furthermore, we use a relatively new method, MRM, that has several advantages including minimal shrinkage of variance and less over-estimation (under-estimation) of health utility among people with poor (good) health states. It also provides further information about the properties of the MRM.

Author contributions Conceptualization: YBC, NL, and HLW and GCHK. Data analysis: YBC and HXT. Writing, original draft: YBC and HXT. Writing, critical review and final version: HXT, YBC, NL, and HLW, GCHK.

Data availability The dataset analysed is not publicly available due to IRB restrictions but is available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest All authors declare that they have no potential conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (National University of Singapore Institutional Review Board S17-257E) and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Herdman, M., Gudex, C., Lloyd, A., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research, 20*(10), 1727–1736.
2. Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 Health Survey. *Journal of Clinical Epidemiology, 51*(11), 1115–1128.
3. Horsman, J., Furlong, W., Feeny, D., et al. (2003). The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health and Quality of Life Outcomes, 1,* 54.
4. Drummond, M. F., Sculpher, M. J., Claxton, K., et al. (2015). *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press.
5. Fayers, P. M., & Hays, R. D. (2014). Should linking replace regression when mapping from profile-based measures to preference-based measures? *Value in Health, 17*(2), 261–265.
6. Fayers, P. M., & Machin, D. (2016). *Quality of life: The assessment, analysis and reporting of patient-reported outcomes* (3rd ed.). Oxford: Wiley.
7. National Institue for Health and Care Excellence. (2013). *Guide to the Methods of Technology Appraisal*. London: National Institute for Health and Care Excellence.
8. Longworth, L., & Rowen, D. (2013). Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. *Value in Health, 16*(1), 202–210.
9. Agency for Care Effectiveness. (2018). *Drug Evaluation Methods and Process Guide*. Singapore: Agency for Care Effectiveness, Ministry of Health.
10. Singapore Ministry of Health. (2016). *ElderShield Fast Facts*. Singapore: Ministry of Health.
11. van Exel, N. J., Scholte Op Reimer, W. J., & Koopmanschap, M. A. (2004). Assessment of post-stroke quality of life in cost-effectiveness studies: The usefulness of the Barthel Index and the EuroQoL-5D. *Quality of Life Research, 13*(2), 427–433.
12. Kaambwa, B., Billingham, L., & Bryan, S. (2013). Mapping utility scores from the Barthel index. *European Journal of Health Economics, 14*(2), 231–241.
13. Shah, S., Vanclay, F., & Cooper, B. (1989). Improving the sensitivity of the Barthel Index for stroke rehabilitation. *Journal of Clinical Epidemiology, 42,* 703–709.
14. Shah, S., Cooper, B., & Maas, F. (1992). The Barthel Index and ADL evaluation in stroke rehabilitation in Australia, Japan, the U K and the USA. *Australian Occupational Therapy Journal, 39,* 5–13.
15. Brazier, J., Yang, Y., Tsuchiya, A., & Rowen, D. L. (2010). A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *European Journal of Health Economics, 11,* 215–225.
16. Dakin, H. (2013). Review of studies mapping from quality of life or clinical measures to EQ-5D: An online database. *Health and Quality of Life Outcomes, 11,* 151.
17. Wee, H. L., Yeo, K. K., Chong, K. J., Khoo, E. Y. H., & Cheung, Y. B. (2018). Mean rank, equipercentile and regression mapping of World Health Organization Qualty of Life Brief (WHOQOL-BREF) to EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L) utilties. *Medical Decision Making, 38*(3), 319–333.
18. Cheung, Y. B., Thumboo, J., Gao, F., et al. (2009). Mapping the English and Chinese versions of the Functional Assessment of Cancer Therapy-General to the EQ-5D Utility Index. *Value in Health, 12*(2), 371–376.
19. Cheung, Y. B., Tan, H. X., Wang, V. W., et al. (2019). Mapping the Alzheimer's disease cooperative study—Activities of daily

living inventory to the Health Utility Index Mark III. *Quality of Life Research, 28*(1), 131–139.
20. Cheung, Y. B., Luo, N., Ng, R., & Lee, C. F. (2014). Mapping the Functional Assessment of Cancer Therapy—Breast (FACT-B) to the 5-level EuroQoL Group's 5-dimension questionnaire (EQ-5D-5L) utility index in a Multi-ethnic Asian Population. *Health and Quality of Life Outcomes, 12,* 180.
21. Cheung, Y. B., Tan, L. C. S., Lau, P. N., Au, W. L., & Luo, N. (2008). Mapping the 8-item Parkinson's Disease Questionnaire (PDQ-8) to the EQ-5D Utility Index. *Quality of Life Research, 17,* 1173–1181.
22. Sullivan, P. W., & Ghushchyan, V. (2006). Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. *Medical Decision Making, 26*(4), 401–409.
23. Mpundu-Kaambwa, C., Chen, G., Russo, R., Stevens, K., Petersen, K. D., & Ratcliffe, J. (2017). Mapping CHU9D Utility Scores from the PedsQL™ 4.0 SF-15. *Pharmacoeconomics, 35*(4), 453–467.
24. Chen, G., Tan, J. T., Ng, K., Iezzi, A., & Richardson, J. (2014). Mapping of incontinence quality of life (I-QOL) scores to assessment of quality of life 8D (AQoL-8D) utilities in patients with idiopathic overactive bladder. *Health and Quality of Life Outcomes, 12,* 133.
25. Longworth, L., Yang, Y., Young, T., et al. (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modellling and survey. *Health Technology Assessment, 18*(9), 1–224.
26. Joyce, V. R., Sun, H., Barnett, P. G., Bansback, N., Griffin, S. C., Bayoumi, A. M., et al. (2017). Mapping MOS-HIV to HUI3 and EQ-5D-3L in patients with HIV. *MDM Policy and Practice, 2*(2), 2381468317716440.
27. Lee, C. F., Ng, R., Luo, N., & Cheung, Y. B. (2018). Development of conversion functions mapping the FACT-B total score to the EQ-5D-5L utility value by three linking methods and comparison with the ordinary least square method. *Applied Health Economics and Health Policy, 16*(5), 685–695.
28. Feeny, D., Furlong, W., Torrance, G. W., et al. (2002). Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. *Medical Care, 40*(2), 113–128.
29. Luo, N., Chew, L. H., Fong, K. Y., et al. (2003). A comparison of the EuroQol-5D and the Health Utilities Index mark 3 in patients with rheumatic disease. *Journal of Rheumatology, 30,* 2268–2274.
30. Abdin, E., Chong, S. A., Seow, E., Peh, C. X., Tan, J. H., Liu, J., et al. (2019). A comparison of the reliability and validity of SF-6D, EQ-5D and HUI3 utility measures in patients with schizophrenia and patients with depression in Singapore. *Psychiatry Research, 274,* 400–408.
31. Royston, P., & Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modeling (with discussion). *Applied Statistics, 43,* 429–467.
32. Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Upper Saddle River: Prentice Hall.
33. Singapore Stroke Registry, Ministry of Health. (2015). *Trends in Stroke in Singapore 2005-2015. Singapore Storke Registry Report No. 4*. Singapore: National Registry of Disease Office, Ministry of Health.
34. Versteegh, M. M., Rowen, D., Brazier, J. E., & Stolk, E. A. (2010). Mapping onto EQ-5D for patients in poor health. *Health and Quality of Life Outcomes, 8,* 141.
35. Kong, K. H., & Yang, S. Y. (2006). Health-related quality of life among chronic stroke survivors attending a rehabilitation clinic. *Singapore Medical Journal, 47*(3), 213–218.