Quality of Life Research (2019) 28:131–139
https://doi.org/10.1007/s11136-018-1991-4

Mapping the Alzheimer's Disease Cooperative Study-Activities of Daily Living Inventory to the Health Utility Index Mark III

Yin Bun Cheung[1,2] · Hui Xing Tan[3] · Vivian Wei Wang[4] · Nagaendran Kandiah[5] · Nan Luo[3] · Gerald C. H. Koh[3] · Hwee Lin Wee[3,6]

Accepted: 30 August 2018 / Published online: 1 September 2018

Abstract

Purpose To map the Alzheimer's Disease Cooperative Study—Activities of Daily Living Inventory (ADCS-ADL) to the Health Utility Index Mark III (HUI3) in people living with dementia (PWD) and to compare the performance of five methods for mapping.

Methods A cross-sectional study of 346 dyads of community-dwelling PWD and family caregiver was carried out in Singapore. ADCS-ADL and HUI3 were rated by the family caregivers. Disease severity ratings and Mini Mental State Examination (MMSE) results were retrieved from medical records. A recently proposed mapping method called the Mean Rank Method (MRM) was described and applied, and the results were compared with regression-based mapping, including ordinary least squares, censored least absolute deviation (CLAD), Tobit and response mapping.

Results The MRM produced a mapped utility distribution that closely resembled the observed utility distribution. The standard deviations (SDs) of the observed and MRM-mapped utility were both 0.340, whereas the SDs of the other mapped utilities ranged from 0.243 (response mapping) to 0.283 (CLAD). Regressing the MRM- and CLAD-mapped and observed utility values upon disease severity and MMSE gave similar regression lines (each $P > 0.05$). Regressing the other mapped utility values upon the covariates under- (over-) estimated the utility of good (poor) clinical states. However, regression-based mapping methods gave a better fit at the individual level, as measured by root mean square error, mean absolute error and R^2. K fold cross-validation gave similar results.

Conclusions The MRM is accurate at the group level. The regression-based mapping methods are more accurate for making individual-level prediction. In addition, CLAD also performed reasonably well at the group level.

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

Introduction

Cost–utility analysis is a common form of economic evaluation of health care and medical treatments [1]. The utility information is derived from preference-based measures of patient outcomes or direct evaluation methods [2]. Several

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11136-018-1991-4) contains supplementary material, which is available to authorized users.

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

1 Center for Quantitative Medicine, Duke-NUS Medical School, Level 6, Academia, 20 College Road, Singapore 169856, Singapore

2 Center for Child Health Research, University of Tampere and Tampere University Hospital, Tampere, Finland

3 Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

4 Department of Hospital Management, Fudan University, Shanghai, China

5 Department of Neurology, National Neuroscience Institute, Singapore, Singapore

6 Department of Pharmacy, National University of Singapore, Singapore, Singapore

preference-based measures are available, such as the EuroQoL 5 Dimensions Questionnaire (EQ-5D) [3], Health Utilities Index Mark III (HUI3) [4] and Short Form-6D (SF-6D) [5]. The EQ-5D and HUI3 were, respectively, the most and second most frequently used measure in the National Institute for Health and Care Excellence (NICE) Technology Appraisals [6]. While both EQ-5D and HUI3 measure important aspects of health outcomes, the HUI3 covers eight attributes, including speech and cognition. These attributes may be more relevant to some diseases such as stroke and dementia than others. The HUI3 may allow a richer description of people living with dementia (PWD).

Although utility measures are essential in cost–utility evaluation, such data are not always available. There has been a strong interest in "mapping" descriptive measures of health to utility values [7–9]. This is sometimes referred to as "cross-walking" [10]. If successfully developed, mapping functions can facilitate cost–utility analysis even if only descriptive measurement data are available. Mapping is accepted by NICE [11, 12], and is described in a major textbook on quality of life research [8].

Two reviews of health utility mapping studies show that a variety of source instruments such as various quality of life scales, the Short Form variants (SF-36 and SF-12), the Health Assessment Questionnaire (and its variants) and the Barthel Index have been mapped to utility values [13, 14]. While measures of activities of daily living (ADL) have been relatively less common source instruments in the reviews, some of them have been mapped to utility values, e.g. the Barthel Index and the Health Assessment Questionnaire-Disability Index. ADL assessment is often performed in regular health care in addition to research studies of people living with chronic diseases or disabilities. For example, in Singapore, ADL assessment is required for claims under the ElderShield insurance scheme [15]. ADL data provide an opportunity to enhance understanding of the impact of diseases and health care options.

The Alzheimer's Disease Cooperative Study-Activities of Daily Living Inventory (ADCS-ADL) is an informant/caregiver-administered scale to measure performance of ADL of people living with Alzheimer's disease [16, 17]. It does not offer a utility value for incorporation into cost–utility analysis.

The HUI3 measures eight attributes, namely vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain. Six of the eight attributes are described in terms of levels of ability. Emotion and pain are described differently. The highest and lowest levels of emotion are described as "Happy and interested in life" and "So unhappy that life is not worthwhile", respectively. The highest and lowest levels of pain are described as "Free of pain and discomfort" and "Severe pain that prevents most activities", respectively. In between the two extremes are pain levels that prevent activities to different degrees. Except emotion, all the attributes are closely related to at least one of the ADCS-ADL items [4, 16].

While ADCS-ADL does not directly measure emotion, at least one of its items—"perform a pastime/hobby/game" (item 23)—indirectly reflects emotion to some extent. There is a substantial degree of conceptual overlap between the HUI3 and ADCS-ADL.

Ordinary least square (OLS) regression is the most commonly used mapping method in the health care context so far [13, 14]. A well-known disadvantage of OLS mapping is that it tends to underestimate variability and therefore risks inflating type 1 error in hypothesis testing [7]. A less discussed but no less important disadvantage is that, owing to the replacement of high and low values by their predicted means, OLS mapping tends to underestimate the utility of good health status and overestimate the utility of poor health status, hence under-estimating the association between utility and health states [18]. For example, among cancer patients with Eastern Cooperative Oncology Group (ECOG) performance score 0 (the best), the mean OLS-mapped EQ-5D utility values was lower than the mean of the observed EQ-5D utility values, whereas the opposite bias was seen in patients with poor ECOG performance score [19]. Various attempts have been made to use other regression models to improve mapping accuracy. The three most commonly used alternatives are censored least absolute deviation (CLAD), Tobit and response mapping [14]. However, there has been no consistent evidence to conclude whether any of them may outperform each other [6, 19–23].

A new mapping procedure, called the Mean Rank Method (MRM), was recently proposed [18]. The method is in principle similar to the equipercentile method, which is popular in the field of education [7, 24, 25]. However, it does not require smoothing. It is simple to understand and is more accurate than the OLS and equipercentile mapping methods in reproducing the features of an observed utility distribution and its association patterns with clinical features and health states [18]. But it is not more accurate than OLS in making individual-level predictions. Its performance has been demonstrated in both simulation and an application to map the World Health Organization Quality of Life-BREF (WHOQOL-BREF) to the EQ-5D utility index [18].

The present study is aimed at mapping the ADCS-ADL to the HUI3 among PWD in Singapore. A secondary aim is to characterize the relative performance of the five mapping methods aforementioned.

Methods

Participants

This is a secondary analysis of data drawn from a cross-sectional study of patients and their caregivers who received care at the dementia clinic of the National Neuroscience Institute, Singapore, between August 2013 and July 2015. Ethics approval had been obtained from the National University of Singapore Ethics Committee. The participants were Singapore residents who were diagnosed with dementia based on DSM IV-TR criteria, and their primary informal caregivers. Primary informal caregiver was defined as "the family member or friend (but not a foreign domestic worker) who is most involved in providing care or ensuring provision of care to the patient". This was the definition used in the Survey on Informal Caregiving [26], a national population-based survey in Singapore for community-dwelling older adults. Similar definitions have also been widely used in the literature [27–29].

Measurements

The Health Utilities Index Mark III, or HUI3, was administered separately by caregivers as proxies and by patients. As with previous studies of people with dementia [30–32], health utility measurement using caregiver proxy but not patient self-administration was found valid and reliable in the Singaporean PWD population [33]. Therefore, in this analysis, only the caregiver-administered HUI3 data were used. The HUI3 measures eight attributes. A multiplicative multi-attribute utility function is available to convert the responses to the eight descriptors into a health utility value [34]. It has a possible range of −0.359 to 1, with 1 representing full health, 0 representing health states not better than death and negative value representing health states worse than death.

Clinical data were retrieved from the medical record system, including the ADCS-ADL [16, 17], disease severity and Mini Mental State Examination (MMSE) [35, 36]. The 23-item version of the ADCS-ADL was administered by the caregivers to rate the patients' performance in activities in the previous 4 weeks. ADCS-ADL is an informant/caregiver-administered scale to measure ability to perform ADL [16, 17]. The total score ranges from 0 to 78, with higher scores indicating better functioning. The MMSE is a multi-domain instrument to assess the level of cognitive function in older adults or patients with neurological conditions. Scores on the test items are summed to give a minimum of 0 (the worst) and maximum of 30 (the best). Disease severity was determined by a team of clinicians (including neurologists, psychologists and nurses) based on their clinical interviews with patients and/or caregivers using the Clinical Dementia Rating (CDR) Scale [37]. The CDR uses information from semi-structured interviews to characterize cognitive and functional performance of people with dementia. The scores 0 and 0.5 represent normal and very mild dementia, respectively. All patients in this study had CDR scores 1 (mild), 2 (moderate) or 3 (severe dementia) [37]. They were recoded as 0, 1 and 2 in our regression analysis so that the intercept has an interpretation of the average utility among those with mild disease. The purpose of including MMSE and CDR in the analysis was to assess whether the pattern of association between the utility values obtained from mapping methods and these covariates reflected the pattern of association between the observed HUI3 utilities and the covariates.

Statistical analysis

The ADCS-ADL was mapped to the HUI3 by the MRM and four regression-based methods, including OLS, CLAD, Tobit and response mapping [6, 7, 14]. As compared to OLS, CLAD and Tobit, the response mapping is an indirect method that first predicts the probabilities of response levels in each of the eight HUI dimensions and then maps these responses to the utility values using probability weights [6].

The MRM is conceptually similar to the equipercentile method [18]. The basic idea of the equipercentile method is that x and y are considered equivalent if $F(x) = P(X \le x) = P(Y \le y) = G(y)$, where $F(x)$ and $G(y)$ are the cumulative distribution functions (CDFs) of variables X and Y, respectively [7, 24, 25]. While the equipercentile approach is conceptually attractive, a major disadvantage is that the solution exists only if there are no tied values. Ties are common in health and utility measurements due to the discrete nature of some measurement scales. In this case, smoothing of the CDFs is needed before the equipercentile mapping can be performed [7, 25]. Smoothing is in general not a simple task. In particular, one of the difficulties in the practice of smoothing is the handling of boundary effects. Health utility data often show a strong ceiling effect [22]. This affects the accuracy of equipercentile mapping and forms a major barrier to using this mapping approach in the health care context [7, 25].

The MRM circumvents these complications by using a simple procedure to handle ties [18]. The procedure is as follows:

(1) The ADCS-ADL scores in a sample are sorted in ascending order, and a rank is assigned to each value. For tied values, the mean of ranks is assigned.
(2) The HUI3 values in a sample are sorted in ascending order, and a rank is assigned to each value. For tied values, the ranking is arbitrary.
(3) Each unique ADCS-ADL score is mapped to the HUI3 value that has the same rank.
(4) In the case of N tied ADCS-ADL scores, they are mapped to the mean of the N consecutive HUI3 values whose mean of ranks equals the mean rank of the N tied ADCS-ADL scores.

For OLS mapping and response mapping by multinomial logistic model, the fractional polynomial (FP) approach with up to two power terms was used to model the possibly non-linear relation between ADCS-ADL scores and HUI3 utilities or response levels, respectively [38]. The deviance difference and Akaike Information Criterion (AIC) were used to guide the two model selection, respectively. For CLAD and Tobit, FP software is not available. The natural polynomials with up to three power terms and the FP identified from OLS were applied to CLAD and Tobit models. The CLAD model with the largest pseudo R-squared (R^2) and the Tobit model with the smallest AIC were selected. The bootstrap method with 1000 replicates was used to estimate standard error of prediction, since the MRM, CLAD and response mapping have no analytical standard error of prediction.

All the data were used in the development of the mapping functions. We used K fold cross-validation, with $K = 10$, to validate the mapping functions [39].

Some popular evaluation criteria inherently favour certain mapping methods; for example, mean squared error and root mean squared error (RMSE) by definition favour OLS mapping. Thus, multiple criteria were used to achieve a broad overview, including evaluation at both the group and individual levels. Firstly, we compared the distribution of the observed and mapped utilities, including percentiles, mean and standard deviation (SD). Secondly, we examined whether the mapped utilities could be used to produce the association patterns between the observed utilities and covariates, including disease severity and MMSE. The analysis began with assessing the gradients between the observed utilities and disease severity and MMSE. For disease severity, we compared a model that treated the 3 levels as categories (2 degrees of freedom) versus a model that assumed a linear trend across the 3 levels (1 degree of freedom) using the likelihood ratio test. For MMSE, we used the FP approach to identify an appropriate form of gradient. Based on the forms identified from the observed data, Zellner's Seemingly Unrelated Regression approach was used to compare the regression analysis results between the observed and mapped utilities [40]. This allows formal comparison of regression coefficients between models. Thirdly, we calculated measures of individual-level agreement between the mapped and observed utility values, including RMSE, mean absolute error (MAE), intra-class correlation (ICC) and the R^2 (square of Pearson's correlation coefficient) in predicting the observed utilities by the mapped utilities.

Results

Descriptive summary

A total of 367 patient-caregiver dyads were approached, among whom 21 dyads declined to participate. Hence, 346 pairs of patient and primary informal caregiver were included in this analysis. Among the 346 patients, MMSE was missing on two patients. There was no other missing values in the variables used. The characteristics of the patients and caregivers are shown in Table 1. The Spearman's correlation coefficient between ADCS-ADL and HUI3 utility was 0.66. Further description of the patients by disease severity is included in Online Supplementary Material S1.

Development of mapping functions

For OLS, a 2-degree FP gave a better fit than a linear trend and a 1-degree FP (each $P < 0.05$). The estimated OLS mapping equation using 2-degree FP was

$$\begin{aligned}\text{HUI3 utility} &= -0.2112 + 0.1505 \times \ln(\text{ADCS-ADL} + 1)\\ &\quad + 7.02 \times 10^{-7} \times (\text{ADCS-ADL} + 1)^3.\end{aligned}$$

When the same FP was applied to CLAD and Tobit, it gave higher pseudo R^2 and smaller AIC, respectively, than the use of natural polynomials with up to three power terms. In the multinomial logistic models for response mapping, fractional and natural polynomials either did not converge or did not fit better than ADL as a linear predictor in 5 out of 8 HUI attributes. ADL as a linear predictor was used in the multinomial logistic modelling. Details of the regression coefficients and their standard errors are available in Online Supplementary Material S2.

The MRM and regression-based mapping table for converting ADCS-ADL scores 0–78 to HUI3 utility values and the bootstrap standard errors of the predictions are provided in Online Supplemental Material S3 as an electronic spreadsheet. Figure 1 plots the observed HUI3 values and describes the MRM, OLS, CLAD and response mapping functions in relation to ADCS-ADL. The Tobit function was very similar to that of OLS and therefore was not included in the figure to avoid confusion. For ADCS-ADL scores below approximately 55, the regression-based mapping functions were higher than the MRM mapping function, and vice versa. Only the MRM function reached the full health utility level (1). Among the regression-based functions, the CLAD reached higher than the others despite not reaching the full health utility level.

Table 2 shows the details of the utility distributions. The mean of MRM-, OLS- and Tobit-mapped utilities agreed with the observed value, 0.416. CLAD and response mapping over- and underestimated the mean, respectively. The SD of the MRM-mapped utility also agreed with that of the observed values, 0.340. The other methods underestimated the SD by 17% (CLAD) to 29% (response mapping). All regression-based methods overestimated the 10th and 25th percentiles and underestimated the 75th and 90th percentiles.

Table 1 Participant characteristics ($N=346$)

Persons	Characteristics	N or mean*	% or SD*
Patient	Age (years)	73.7	9.8
	Gender		
	Male	128	37.0%
	Female	218	63.0%
	Ethnicity		
	Chinese	312	90.2%
	Malay	12	3.5%
	Indian	15	4.3%
	Others	7	2.0%
	Disease duration (years)	2.2	1.8
	Disease severity		
	Mild	206	59.5%
	Moderate	78	22.5%
	Severe	62	17.9%
	Type of dementia		
	Alzheimer's disease	248	71.7%
	Vascular dementia	46	13.3%
	Others	52	15.0%
	MMSE	16.6	6.9
	ADCS-ADL	43.6	22.9
	HUI3	0.416	0.340
Caregiver	Relationship to patient		
	Spouse	131	37.9%
	Child	205	59.2%
	Sibling/parent	10	2.9%
	Gender		
	Male	118	34.1%
	Female	228	65.9%

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

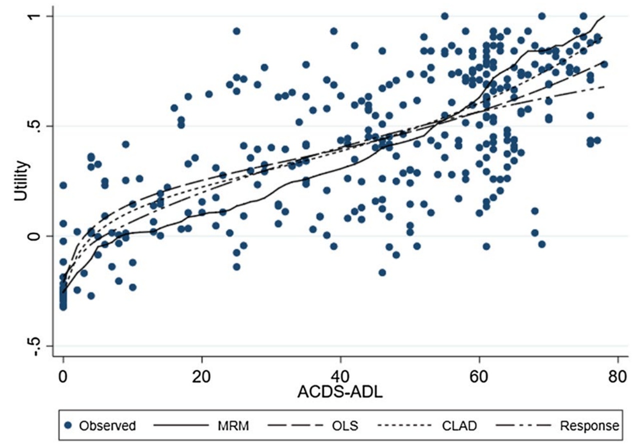

Fig. 1 Scatter-plot of observed HUI3 utility values and mapping functions derived from MRM, OLS, CLAD and response mapping in people living with dementia ($N=346$)

In contrast, the MRM method closely mirrored the observed utilities in terms of all these percentiles. The MRM-mapped utility distribution reached the full health utility, 1. The other methods at most reached 0.909 (CLAD). Among the four regression-based methods, CLAD provided a utility distribution that was relatively similar to the observed distribution in terms of SD and percentiles.

In the observed HUI3 data, a linear trend across three levels of disease severity showed no significant difference in likelihood as compared to modelling disease severity as a categorical variable ($P=0.227$); a linear trend across all MMSE scores fitted equally well as compared to 2-degree FP ($P=0.967$). Hence the linear trends of observed/mapped utilities in relation to the covariates were assessed. Table 3 compares the regression findings using the Zellner's Seemingly Related Regression method. The MRM-mapped, CLAD-mapped and observed utilities gave similar trends of mean utility in relation to disease severity (each $P>0.10$). In contrast, the OLS, Tobit and response mapping methods showed trends different from the observed (each $P<0.01$). In particular, they underestimated the intercept (each $P<0.05$), i.e. the mean utility of those with mild disease, and they underestimated the gradient of utility in relation to disease severity (each $P<0.001$). For MMSE, the MRM-mapped, CLAD-mapped and observed utilities gave similar trends of mean utility in relation to MMSE (each $P>0.10$). In contrast, the OLS, Tobit and response mapping methods produced different trends (each $P<0.01$). In particular, they overestimated the intercept (each $P<0.01$), i.e. the mean utility of those at the worst possible MMSE level, and they underestimated the gradient of utility in relation to MMSE (each $P<0.001$).

Table 4 shows the measures of agreement between the observed and mapped utilities at the individual level. OLS and Tobit had the smallest RMSE and largest R^2 in predicting the observed values. CLAD had the strongest agreement with the observed values in terms of MAE and ICC. The MRM had the largest RMSE, MAE and smallest R^2, but the second highest ICC.

Cross-validation

The results are shown in the lower panels of Tables 2, 3 and 4. The findings are mostly very similar to those in the whole study sample. A difference is that, in the whole study sample, the MRM-mapped and observed utilities gave similar trends of mean utility in relation to MMSE ($P=0.169$). But in the cross-validation sample, they gave somewhat different trends ($P=0.072$).

Table 2 Distribution of observed and mapped HUI3 utilities in the whole study sample ($N=346$) and in 10-fold cross-validation sample

Samples	Utility	Mean	SD	Min	P10	P25	P50	P75	P90	Max
Whole study	Observed	0.416	0.340	−0.323	−0.047	0.145	0.431	0.710	0.842	1.000
	MRM-mapped	0.416	0.340	−0.256	−0.048	0.145	0.427	0.719	0.842	1.000
	OLS-mapped	0.416	0.244	−0.211	0.059	0.299	0.465	0.588	0.669	0.793
	CLAD-mapped	0.429	0.283	−0.257	0.020	0.272	0.472	0.632	0.741	0.909
	Tobit-mapped	0.416	0.245	−0.211	0.058	0.298	0.466	0.590	0.672	0.797
	Response-mapped	0.396	0.243	−0.184	−0.015	0.265	0.481	0.579	0.625	0.678
Cross-validation	Observed	0.416	0.340	−0.323	−0.047	0.145	0.431	0.710	0.842	1.000
	MRM-mapped	0.405	0.347	−0.323	−0.047	0.145	0.423	0.692	0.842	1.000
	OLS-mapped	0.415	0.244	−0.229	0.057	0.301	0.464	0.584	0.671	0.800
	CLAD-mapped	0.432	0.281	−0.263	0.029	0.286	0.477	0.629	0.736	0.922
	Tobit-mapped	0.416	0.245	−0.229	0.056	0.301	0.464	0.586	0.673	0.805
	Response-mapped	0.396	0.243	−0.188	0.023	0.268	0.479	0.576	0.626	0.680

P percentile

Table 3 Regression analysis of observed and mapped HUI3 utilities in relation to disease severity and MMSE in the whole study sample ($N=346$) and in 10-fold cross-validation sample

Samples	Covariates	Utilities	Intercepts	*P**	Slopes	*P**	Model *P***
Whole study	Disease	Observed	0.574		−0.272		
	Severity	MRM-mapped	0.586	0.498	−0.292	0.260	0.530
		OLS-mapped	0.540	0.030	−0.214	<0.001	0.002
		CLAD-mapped	0.573	0.939	−0.247	0.139	0.199
		Tobit-mapped	0.542	0.038	−0.215	<0.001	0.002
		Response-mapped	0.522	0.001	−0.215	<0.001	0.001
Whole study	MMSE	Observed	−0.053		0.028		
		MRM-mapped	−0.120	0.071	0.032	0.060	0.169
		OLS-mapped	0.052	0.001	0.022	<0.001	0.001
		CLAD-mapped	−0.002	0.125	0.026	0.189	0.298
		Tobit-mapped	0.051	0.001	0.022	<0.001	0.002
		Response-mapped	0.040	0.004	0.021	<0.001	<0.001
Cross-validation	Disease	Observed	0.574		−0.272		
	Severity	MRM-mapped	0.579	0.774	−0.299	0.133	0.241
		OLS-mapped	0.540	0.029	−0.213	<0.001	0.002
		CLAD-mapped	0.575	0.949	−0.245	0.112	0.125
		Tobit-mapped	0.541	0.036	−0.214	<0.001	0.002
		Response-mapped	0.521	0.001	−0.214	<0.001	0.001
Cross-validation	MMSE	Observed	−0.053		0.028		
		MRM-mapped	−0.138	0.023	0.033	0.037	0.072
		OLS-mapped	0.053	0.001	0.022	<0.001	0.001
		CLAD-mapped	0.006	0.074	0.026	0.139	0.185
		Tobit-mapped	0.052	0.001	0.022	<0.001	0.002
		Response-mapped	0.041	0.004	0.021	<0.001	<0.001

*Test of difference in intercept or slope between mapped and observed data, **joint test of differences in intercept and slope between mapped and observed data

Discussion

The ADCS-ADL is a useful tool for the assessment of PWD. The HUI3 is the second most frequently used preference-based measure in NICE Technology Appraisals [6]. They have substantial conceptual overlap. Mapping the ADCS-ADL to the HUI3 facilitates the evaluation of care and treatments when utility data or preference-based measures have not been directly collected.

Table 4 Root mean squared errors (RMSE), mean absolute errors (MAEs), intra-class correlation (ICC) coefficient and R-squared (R^2) of mapped HUI3 utilities compared to observed utilities in the whole study sample ($N=346$) and in 10-fold cross-validation sample

Samples	Utilities	RMSE	MAE	ICC	R^2
Whole study	MRM-mapped	0.265	0.203	0.697	0.484
	OLS-mapped	0.237	0.191	0.682	0.517
	CLAD-mapped	0.240	0.187	0.706	0.514
	Tobit-mapped	0.237	0.191	0.683	0.515
	Response-mapped	0.239	0.196	0.672	0.505
Cross-validation	MRM-mapped	0.266	0.206	0.699	0.489
	OLS-mapped	0.237	0.192	0.679	0.512
	CLAD-mapped	0.241	0.190	0.702	0.511
	Tobit-mapped	0.237	0.192	0.680	0.512
	Response-mapped	0.240	0.197	0.669	0.501

Mapping by the OLS method is known to have several disadvantages [7, 18]. Other regression-based mapping methods have been investigated but they did not consistently outperform the OLS method [6, 19–23]. A well-known feature of OLS mapping is that $r(\hat{y}_{\text{OLS}}, y) = r(x, y)$, where x is the predictor, y is the utility, r is the Pearson's correlation coefficient and $\hat{y}_{\text{OLS}}$ is the OLS-mapped utility. In contrast, MRM is rank-based. If there are no ties, the rank of each MRM-mapped utility, $\hat{y}_{\text{MRM}}$, is identical to the rank of its corresponding predictor value. Consequently, MRM has a feature of $\rho(\hat{y}_{\text{MRM}}, y) = \rho(x, y)$, where rho is the Spearman rank correlation coefficient. This equality can be upset by ties in utility values but not by ties in the predictor values [18].

Our findings corroborated previous simulation and empirical data that the MRM caused little shrinkage of variance [18, 41]. Furthermore, the MRM and CLAD approximated well the association pattern of observed utility levels in relation to dementia disease severity and MMSE, whereas the OLS, Tobit and response mapping showed a biased pattern. In cost–utility analysis, healthcare recipients under each intervention (treatment or healthcare model) may be classified into ordered health states, H_k ($k=1, 2,\ldots,K$), with H_1 and H_K being the most and least desirable health states, respectively. For example, H_1, H_2 and H_3 ($K=3$) may be mild, moderate and severe dementia according to the CDR. An intervention may aim to shift the distribution of health states towards H_1, for example, to slow down progression towards severe disease. The gain in quality-adjusted life years (QALYs) generated by an intervention of interest ($j=2$) versus its comparator ($j=1$) is estimated by

$$\text{Gain in QALY} = \sum_{k=1}^{K} P_{2,k}\mu_k y_k - \sum_{k=1}^{K} P_{1,k}\mu_k y_k,$$

where $P_{j,k}$ is the proportion of people under treatment j who have health state k $\left(\sum_{k=1}^{K} P_{j,k} = 1\right)$; μ_k is the mean utility for health state k and y_k is the life years in health state k [18]. Although OLS and Tobit were accurate in estimating the overall mean utility, μ, MRM and CLAD were more accurate in estimating the means of utility in different health states, μ_k, as indicated by their regression lines in relation to disease severity and MMSE that were comparable with those estimated from the observed utility data. Therefore, MRM and CLAD would likely be more accurate in the estimation of the difference in mean utility between health states (e.g. $\mu_1 - \mu_2$) and gain in QALY than the other methods. The shallower slopes in the regression models of the other mapped utilities in relation to disease severity and MMSE indicated that they have a tendency to underestimate the difference in mean utility between health states and consequently underestimate the gain in QALY. That said, in the present study, we only examined two health measures, disease severity and MMSE. While they shed light on the features of the various mapping methods and corroborated previous simulation results [18], these findings may not be generalizable to all other measures of health states. Further studies on other health measures will be useful.

At the individual level, all regression-based methods gave smaller RMSE and MAE and higher R^2 than the MRM. This is not surprising because the MRM equates two distributions instead of operating with pairs of data points. In fact, the MRM does not even require both X and Y data to be available from the same individuals. For example, if a sample is randomly split into two survey modules, one with X and the other with Y, the MRM still works [18]. The contrast of the findings demonstrates the importance of choosing mapping methods in relation to specific research aims.

MRM is a recently proposed mapping method. In the absence of ties, it is identical to the equipercentile method. In the presence of ties, it simply maps the tied predictor scores to the mean of the corresponding target scores. Based on the findings so far, we expect that the MRM will be useful in describing distribution and association patterns and in cost–utility analysis. However, it has larger prediction error than regression-based methods in making individual-level prediction.

Among the four regression-based methods, CLAD performed better in terms of causing less shrinkage of SD, reaching a higher maximum, producing association patterns in relation to disease severity and MMSE that were similar to the observed data pattern, and having smaller MAE and higher ICC. Its performance appeared to be fairly similar to that of the MRM in terms of the group-level evaluation criteria used. Previous studies have not consistently found CLAD to perform better than other regression-based mapping methods. However, the evaluations were mainly based on individual-level measures like MAE, MSE and RMSE [6, 20, 21]. Extending the evaluation criteria to include the group-level measures used here may shed new lights on the relative performance of CLAD and the other methods.

There are several limitations in this study. The sample size was relatively small for this type of studies. As compared to the studies reviewed in [13], it stood at the 32nd percentile. The validation has only used internal data. The study involved only community-dwelling people with dementia in Singapore. Presently, the scope of application of the mapping functions is limited to this population. Further validation of the mapping functions in larger samples and in broader geographical contexts and institutional care settings would help to further evaluate their performance and the generalizability of the present findings.

Author contributions VWW, NK and HLW designed and conducted the cross-sectional study of dementia patients and caregivers. YBC and HLW conceived this specific aim for mapping ADL inventory to health utilities. YBC, HLW, NL and GCHK contributed to the development of the analysis strategy. YBC and HXT implemented the statistical analysis. YBC wrote the first draft of the article. All the authors critically reviewed the article and agreed with the submission.

Compliance with ethical standards

Conflict of interest All authors declare that they have no 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 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. Drummond, M. F., Sculpher, M. J., Claxton, K., et al. (2015). *Methods for the economic evaluation of health care programmes*. Oxford: Oxford University Press.
2. Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: The QALY and utilities. *British Medical Bulletin, 96*, 5–21.
3. 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.
4. 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.
5. 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.
6. 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 modelling and survey. *Health Technology Assessment*. https://doi.org/10.3310/hta18090.
7. 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.
8. Fayers, P. M., & Machin, D. (2016). *Quality of life: The assessment, analysis and reporting of patient-reported outcomes* (3rd ed.). Oxford: Wiley.
9. Whately-Smith, C., Watkins, C., Mann, H., Fletcher, C., & Ducournau, P. (2014). Utility values in health technology assessments: A statistician's perspective. *Pharmaceutical Statistics, 13*(3), 184–195.
10. Crott, R. (2014). Mapping algorithms from QLQ-C30 to EQ-5D utilities: No firm ground to stand on yet. *Expert Review of Pharmacoeconomics and Outcomes Research, 14*(4), 569–576.
11. National Institute for Health and Care Excellence. (2013). *Guide to the methods of technology appraisal*. London: National Institute for Health and Care Excellence.
12. 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.
13. 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.
14. 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.
15. Singapore Ministry of Health. (2016). *ElderShield fast facts*. Singapore: Ministry of Health. http://www.eldershield.sg.
16. Galasko, D., Bennett, D., Sano, M., et al. (1997). An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Disease and Associated Disorders, 11*(Suppl. 2), S33–S39.
17. Robert, P., Ferris, S., Gauthier, S., et al. (2010). Review of Alzheimer's disease scales: Is there a need for a new multi-domain scale for therapy evaluation in medical practice? *Alzheimer's Research and Therapy, 26*(2), 24.
18. 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 Quality of Life Brief (WHOQOL-BREF) to EuroQoL 5 Dimensions 5 Levels (EQ-5D–5L) utilities. *Medical Decision Making, 38*(3), 319–333.
19. 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.
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. Gray, A. M., Rivero-Arias, O., & Clarke, P. M. (2006). Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Medical Decision Making, 26*(1), 18–29.
22. Huang, I. C., Frangakis, C., Atkinson, M. J., et al. (2008). Addressing ceiling effects in health status measures: A comparison of techniques applied to measures for people with HIV disease. *Health Services Research, 43*, 327–339.
23. 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.
24. Dorans, N. J. (2007). Linking scores from multiple health outcome instruments. *Quality of Life Research, 16*(Suppl. 1), 85–94.

25. Holland, P. W., & Thayer, D. T. (1989). *The kernel method of equating score distributions*. Washington, DC: Educational Testing Service.
26. Chan, A., Ostbye, T., Malhotra, R., & Hu, A. J. (2012). The survey on informal caregiving: The summary report for MCYS. Retrieved June 1, 2013, from https://app.msf.gov.sg/Portals/0/Informal%20Caregiver%20Survey%20Summary%20Report%20(upload).pdf.
27. Contador, I., Fernandez-Calvo, B., Palenzuela, D. L., Migueis, S., & Ramos, F. (2012). Prediction of burden in family caregivers of patients with dementia: A perspective of optimism based on generalized expectancies of control. *Aging and Mental Health, 16*(6), 675–682.
28. Haro, J. M., Kahle-Wrobleski, K., Bruno, G., et al. (2014). Analysis of burden in caregivers of people with Alzheimer's disease using self-report and supervision hours. *Journal of Nutrition, Health and Aging, 18*(7), 677–684.
29. Reed, C., Belger, M., Dell'agnello, G., et al. (2014). Caregiver burden in Alzheimer's disease: Differential associations in adult–child and spousal caregivers in the GERAS observational study. *Dementia and Geriatric Cognitive Disorders Extra, 4*(1), 51–64.
30. Bhattacharya, S., Vogel, A., Hansen, M. L., et al. (2010). Generic and disease-specific measures of quality of life in patients with mild Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders Extra, 30*, 327–333.
31. Coucill, W., Bryan, S., Bentham, P., Buckley, A., & Laight, A. (2001). EQ-5D in patients with dementia: An investigation of inter-rater agreement. *Medical Care, 39*(8), 760–771.
32. Karlawish, J. H., Zbrozek, A., Kinosian, B., Gregory, A., Ferguson, A., & Glick, H. A. (2008). Preference-based quality of life in patients with Alzheimer's disease. *Alzheimer's and Dementia, 4*(3), 193–202.
33. Wang, W. (2016). Economic and health related quality of life outcomes among community-dwelling dementia patients in Singapore. PhD Dissertation, National University of Singapore, Singapore.
34. 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.
35. Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research 1975, 12*(3), 189–198.
36. Feng, L., Chong, M. S., Lim, W. S., & Ng, T. P. The Modified Mini-Mental State Examination test: Normative data for Singapore Chinese older adults and its performance in detecting early cognitive impairment. *Singapore Medical Journal, 53*(7), 458–462.
37. Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology, 43*(11), 2412–2414.
38. Royston, P., & Altman, D. G. (1994). Regression using fractional polynomials of continuous covariates: Parsimonious parametric modeling (with discussion). *Applied Statistics, 43*, 429–467.
39. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning, with applications in R*. New York: Springer.
40. Greene, W. H. (2012). *Econometric analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
41. 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*. https://doi.org/10.1007/s40258-018-0404-8 (E-pub ahead of print).