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

Hwee Lin Wee, PhD[1,2], Khung Keong Yeo, MBBS, ABIM[3,4], Kok Joon Chong, BSc (Hons)[5], Eric Yin Hao Khoo, MD, FRCP[5,6], Yin Bun Cheung, PhD [7,8]

1. Department of Pharmacy, Faculty of Science, National University of Singapore, Singapore
2. Saw Swee Hock School of Public Health, National University of Singapore, Singapore
3. National Heart Centre Singapore, Singapore
4. Duke-NUS Medical School, Singapore.
5. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
6. Division of Endocrinology, University Medicine Cluster, National University Hospital, Singapore
7. Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore
8. Tampere Center for Child Health Research, University of Tampere and Tampere University Hospital, Finland

Corresponding Author:

Dr WEE Hwee Lin, Assistant Professor, Department of Pharmacy, Faculty of Science, & Saw Swee Hock School of Public Health, National University of Singapore, Singapore

E-mail: phawhl@nus.edu.sg

Abstract (Limit: 275 words; Word count: 271 words)

Background: Existing methods to link preference-based and profile-based health-related quality of life (HRQoL) questionnaires have their limitations. Hence, we developed a new mapping method (the mean rank method, MRM) and applied it to map the World Health Organization Quality of Life Questionnaire (WHOQOL-BREF) to the EuroQoL 5 Dimensions 5 Levels (EQ-5D-5L). We then compared the new MRM with current methods, i.e. regression-mapped (OLS method) and equipercentile method (EPM).

Methods: Singapore residents, aged ≥21 years, were recruited from the general population and two outpatient clinics in acute care hospitals. Performance of the MRM was evaluated using both simulation and split-sample validation (n=658 in training and n=657 in validation samples). Using the training sample, we derived three sets of mapped EQ-5D-5L utilities based on MRM, OLS and EPM. Using simulation and the validation sample, we compared the performance of the mapping methods in terms of distribution parameters, mean utility by strata, association with health covariates, and prediction errors at the individual-level, among others.

Results: The WHOQOL-BREF Physical Health domain is the only domain significantly associated with EQ-5D-5L utilities. Simulation showed that MRM more accurately reproduced the variance and percentiles of the distribution of the observed utilities than did the OLS method or EPM. OLS tended to under-estimate the mean utility of good health states, overestimate the mean utility of

poor health states and underestimate the association with covariates. An analysis of validation sample gave similar results.

Conclusion: In scenarios similar to the mapping of WHOQOL-BREF to the EQ-5D-5L, the MRM outperformed OLS method and EPM in important - though not all - aspects. The simplicity and reproducibility of the MRM makes it an attractive alternative to current methods.

Introduction

Economic evaluation of drugs, devices and healthcare programmes is increasingly prevalent, driven in large part by the recognition that healthcare resources are limited. Cost-utility analyses (CUA), which evaluate the incremental cost incurred for an incremental unit of quality-adjusted life years (QALY) gained, is the most common form of economic evaluation.(1) QALY is a survival outcome adjusted for the quality of those additional life years gained, with the quality adjustment factor derived from preference-based measures of health-related quality of life (HRQoL). Several preference-based HRQoL measures are available and include the EuroQoL 5 Dimension 5 Levels (EQ-5D-5L),(2) the Short-Form 6 Dimension (SF-6D)(3) and Health Utilities Index III (HUI-3);(4) the EQ-5D-5L is the preferred measure by the National Institute for Health and Clinical Effectiveness (NICE) in the United Kingdom(5) and the HUI-3 used in several major national population health surveys in Canada.(4)

Although preference-based HRQoL measures are essential in economic evaluation, profile-based HRQoL measures remain the predominant - if not only - measures used in clinical trials and clinical studies as they provide richer information about patient's HRQoL and are more sensitive to clinically relevant changes. For instance, the Short-Form 36 version 2 questionnaire, from which the SF-6D was derived, provides scores on eight aspects of HRQoL including physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE) and mental

health (MH). Hence, there has been strong interest to bring the two types of measures (profile-based and preference-based) together using processes such as mapping or linking (6-8). In a review published in 2010,(9) a total of 30 studies were identified with EQ-5D being the most widely used preference-based HRQoL measure and the Short Form-12 (SF-12) and SF-36 being the most widely-used profile-based HRQoL measure in the mapping studies. Interestingly, to the best of our knowledge, there has not been any study that maps the WHOQOL-BREF to a preference-based measure, despite WHOQOL-BREF being a widely used questionnaire. WHOQOL-BREF(10) is an abbreviated version of the WHOQOL-100(11) developed by the World Health Organization at 18 international field centers simultaneously. It has been used in more than 100 studies worldwide. Hence, the first aim of this study is to map the WHOQOL-BREF to EQ-5D-5L as this has tremendous potential for numerous existing and future studies.

Mapping by ordinary least square regression (OLS) is the most common in quality of life and health utility studies. A major disadvantage of OLS-based mapping is that it tends to underestimate variability and risks inflating type 1 error in hypothesis testing.(6) Equipercentile mapping (EPM) has been popular in educational settings.(6, 12, 13) The basic idea is that x and y are considered equivalent if $F(x) = P(X \leq x) = P(Y \leq y) = G(y)$, where $F(x)$ and $G(y)$ are the cumulative distribution functions (CDF) of X and Y, respectively. While the EPM is conceptually attractive, a major disadvantage is that the solution exists only if

the CDFs are continuous and strictly increasing. For discrete data, which is very common in HRQoL measurements, smoothing of the CDFs is needed before the EPM can be performed.(6) However, smoothing is, in general, not a trivial task. One of the difficulties in the practice of smoothing is boundary effects. In particular, the EQ-5D-5L utilities often show a heavy ceiling effect,(14) which accentuates the problem of boundary effects. Hence, the second aim of this study is to propose a new method based on ranks, and compare its performance with the two mapping methods aforementioned. This method is conceptually similar to the EPM but the practice does not involve smoothing and is very simple to understand and implement. We have organized this manuscript according to the MAPS statement.(15)

Methods

Participants and Study Design

This study comprises participants from the general population as well as samples from two clinics. Ethics approval for this study was obtained from the National Healthcare Group Domain Specific Review Board (Ref. 2013/00747) as well as the SingHealth Centralised Institutional Review Board (Ref. 2015/2041). The clinic samples were chosen to enrich the dataset, covering a wider spread of health status and enhancing the quality of mapping.

Singapore is a multi-ethnic society mainly consisting of Chinese, Malay and Indian people. English is the lingua franca. The general population participants were recruited by using a multi-stage cluster sampling using postcodes as the primary sampling unit (PSU) followed by the selection of household and then the selection of respondents. Three call attempts (1st attempt and 2 call backs) were made on different days and at different times of the week. Only one participant per household was selected. Participants were selected based on a pre-specified quota for language of interviews within each ethnicity interlocked with age and gender. The face-to-face interviews were conducted in the participants' home between October 2014 to January 2015.

The clinic participants were drawn from two separate studies in outpatient clinics in the National Heart Center Singapore (NHCS) and the Division of Endocrinology in the National University Hospital (NUH). Recruitment was conducted by research assistants via convenience sampling: patients were approached in the clinics while waiting to see the doctor. In both clinical samples, patients must respond positively to either of these questions: "Have you ever been told by a doctor previously that you have at least one of the following: blockage of the arteries to your heart, stroke, heart attack, peripheral arterial disease, or transient ischaemic attack, or kidney disease?". and "Have you ever had at least one of the following: heart bypass operation, stent insertion or brain surgery for stroke?". Patients with recent acute myocardial infarction (STEMI), hemodynamic instability or gestational diabetes were

excluded. Interviews were carried out between March 2015 and February 2016.

For the general population and clinic samples, to be eligible, the participant must be a Singapore Resident (include Singapore Citizens and Permanent Residents) aged 21 years and above who speaks English, Chinese (Mandarin) or Malay. Participants who speak only Tamil were excluded as the questionnaires are not available in Tamil. All participants read and signed the written informed consent form prior to commencement of the interviews.

Questionnaires

A total of three HRQoL questionnaires were used (EQ-5D-5L, WHOQOL-BREF and SF-36v2), in addition to a socioeconomic and clinical questionnaire, where information such as age, gender, ethnicity and self-reported medical conditions were captured.

WHOQOL-BREF

The WHOQOL-BREF is a 26-item questionnaire which includes one item from each of the 24 facets contained in the WHOQOL-100 and two additional items on overall quality of life and general health. The 24 items are organized into 4 domains, namely Physical Health, Psychological, Social Relationships and

Environment. Three negatively phrased items were reversed scored. According to the users manual, domain scores were computed by taking the mean of the scores of the items that constitute the domain and multiplied by 4, so that the scores are directly comparable with those derived from WHOQOL-100. However, in our analyses, the raw domain scores (i.e. sum of item scores within domain) were used. Unlike the transformed domain scores, the raw scores had the advantage of equal-interval, which would make the mapping easier, as smoothing of the CDF was needed for one of the mapping methods. We used the mean substitution method to replace missing values provided that there was no more than one missing value per domain, which is the official WHOQOL-BREF approach for handling missing data.(16) That is, missing item scores were replaced with the mean of the non-missing item scores in the same domain; other ways of handling missing values were not considered because our aim was to map the WHOQOL-BREF scores to the EQ-5D-5L utility index. The mean substitution is an integral part of the WHOQOL-BREF score. To use an alternative method for handling the missing values would mean generating a score different from the WHOQOL-BREF, which was not our intention.

EQ-5D-5L

The EQ-5D-5L comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression with 5 response options for each dimension (no, slight, moderate, severe, extreme problems/unable).

Participants indicate their health status for the day by selecting one response option per dimension, giving rise to a 5-digit health profile. This, in turn, is linked with a utility score. We used the Japanese value set from the EuroQoL Group's crosswalk project as a Singapore dataset is not available.(17) The Japanese value set has a possible range of –0.111 to 1.(18)

SF-36v2

The SF-36v2 is a 36-item profile-based generic HRQoL questionnaire that assesses 8 domains of HRQoL, as described above. Scores for the 8-scale dimensions were standardized to the Singapore adult population (mean 50, standard deviation 10) with higher scores indicating better HRQoL; this makes it possible to meaningfully compare scores across domains.(19) Missing data were replaced by mean substitution.

Statistical Analyses

Participants who had non-missing values in the four WHOQOL-BREF domains were randomly split to form a training sample and a validation sample, with approximately equal number of participants in each sample.

Exploratory Data Analysis

As mapping should be performed only if the two instruments measure the same constructs, we performed least squares regression to assess whether all or only some of the four WHOQOL-BREF domain scores were associated with EQ-5D-5L utilities. Domain scores not associated with EQ-5D-5L utilities were excluded from the mapping exercise.

Modelling Approaches and Estimation Methods

Three different mapping methods were employed. First, we proposed a new method referred to as the mean rank method (MRM).

Mean Rank Method for Mapping Health Utility

(1) X is the predictor variable and its values are ranked from smallest (=1) to largest (=N). For tied values, mean of ranks is assigned. For unique (non-tied) values, mean rank is the same as the rank.

(2) Y is the health utility variable and its values are ranked from smallest (=1) to largest (=N). For tied values, the ranking among the set of tied values is arbitrary. The index i refers to the i^{th} row of the sorted (x, y) data.

(3) For each unique x value, x is mapped to the y value with the same rank.

(4) For n_k tied x values at the k^{th} level of unique values in X $(k = 1,2, \dots, K)$, x is mapped to the mean of the n_k consecutive y values whose mean of ranks equals the mean ranks of the tied x values. The vector of mapped values is denoted by $\hat{y}_{MRM}$.

The spirit of the proposed method is like that of the EPM, with ties in X handled with a simple assignment of the mean of the y values with comparable ranks. Unlike the EPM, the proposed method does not achieve symmetry; it is like the OLS method in this regard. In the absence of ties in X, the CDF of the mapped values will be identical to the CDF of the observed Y, regardless of the presence or absence of ties in Y. For tied values in X, the CDF jumps at the mean of the set of y values whose mean of ranks equals the mean ranks of the tied x values. The proposed method involves minimal modelling other than assigning the mean of consecutive y values to their corresponding tied values in X. In the absence of ties in Y, the rank of each $\hat{y}_{MRM}$ value is identical to the rank of its corresponding x value. As such, although there is no direct modelling of the association between variables, the proposed method has a feature of $rho(\hat{y}_{MRM}, y) = rho(x, y)$, where *rho* is the Spearman correlation coefficient, except when there are ties in Y. This is similar to the feature of $r(\hat{y}_{OLS}, y) = r(x, y)$ in OLS-based mapping, where *r* is the Pearson's correlation coefficient and $\hat{y}_{OLS}$ is the mapped utility variable using the OLS method.

The proposed method does not necessarily require X and Y to be paired data from the same individual. For example, suppose that for the purpose of

reducing burden on respondents, individuals are randomized to complete different assessment packages that contain, among other instruments, either X or Y. The proposed method is still usable even though the individual did not complete both X and Y.

As shown in details in Online Appendix 1, the means of the observed and mapped values are identical, but the MRM underestimates the true variance by $\sum_{i=1}^{N} e_i^2/N \geq 0$, where the subscript i denotes the i^{th} row of the sorted (x, y) data, as mentioned above. In the context of OLS-based mapping, Chan et al. (2014) proposed to use 1/ R^2 from a training dataset as a correction factor to inflate the variance of $\hat{y}_{OLS}$ in future studies to correctly estimate the true variance. Similar to their proposal, in the MRM, an R^2-type of variance adjustment factor, $Var(y_i)/\{Var(y_i) - (\sum_{i=1}^{N} e_i^2/N)\}$, can be obtained from the training dataset for use in future studies. If there is no tie in X, the variance of $\hat{y}_{MRM}$ is identical to that of the observed utilities. In practice, the larger the number of levels a measurement scale has, the smaller the number of ties and the closer the variances of the observed and mapped utilities. As shown in Online Appendix 1, as the number of levels on the predictor scale increases and the amount of tied observations reduces, mapping errors reduce.

Regression and Equipercentile Methods for Mapping Health Utility

Second, we used the OLS method to map the WHOQOL-BREF scores to EQ-5D-5L utilities. Non-linear relation was estimated by the fractional polynomial

method. Third, we used the EPM where kernel smoothing with the Epanechnikov kernel function was used to smooth the CDFs.(20) The pseudo-data method was used to mitigate the boundary effect.(21, 22) Kernel smoothing often defines an "optimal" bandwidth as the bandwidth that minimizes the mean integrated squared error if the data follow a Gaussian distribution. For non-Gaussian data, this "optimal" bandwidth tends to be too wide.(23) We considered this "optimal" bandwidth the upper limit and used half of this value in this article. Varying the bandwidth around this half value and below the "optimal" value gave similar results (details not shown). EPM was implemented by linking the two smoothed CDFs.

Validation Methods

The choice of criteria to compare the performance of mapping methods is not straightforward, as different criteria may favour different methods; e.g. minimum sum of squared errors would favour the OLS method. We proposed that the key evaluation criteria should be based on how well the results of the analyses that used the mapped utilities agree with the results of the analyses that used observed utilities. In general, utilities may be used in at least two ways: (1) to describe a population and (2) to estimate group differences or associations. For example, if utilities are used to describe a population, one would expect the descriptive summary, including mean, variance, median, and 10^{th} and 90^{th} percentiles, of the mapped and observed utilities to be similar. In addition, the proportion with low utility should be similar.

If utilities are used to estimate group differences or associations, then in the regression model, one expects the beta coefficients obtained using the mapped utilities and the observed utilities to be similar. We discuss this in the context of a CUA, which is the typical use of utilities to estimate group differences or associations. In CUA, the incremental cost for gaining an incremental quality-adjusted life year (QALY) with a newer treatment (j=2) over an existing treatment (j=1; usually standard of care) is calculated. Patients under each treatment are classified into ordered health states, H_k (k=1, 2, …, K), according to observable signs or measures, with H_1 and H_K being the most and least desirable health states, respectively. For example, H_1 may be Type 2 diabetes mellitus without complications while H_K is Type 2 diabetes mellitus with end stage renal failure. The goal of the new treatment is to shift the distribution of health states towards H_1 (i.e. slow down disease progression). The gain in QALY with the new treatment (j=2) as compared with the existing treatment (j=1) is calculated by $Gain = \sum_{k=1}^{K} P_{2,k}\mu_k y_k - \sum_{k=1}^{K} P_{1,k}\mu_k y_k$, where $P_{j,k}$ is the proportion of patients under treatment j who have health state k, μ_k is the mean utility for health state k and y_k is the length of time spent in health state k. Hence, the key concern in CUA is that the difference in mean utility between groups $(\mu_k - \mu_{k'})$ must be correctly estimated. Otherwise, the treatment's impact on shifting health states would not be properly valued. As such, the mean mapped utilities by health states needs to be similar to that of utilities obtained directly using the EQ-5D-5L or other utility instruments.

In summary, we considered the key criteria for comparing different mapping methods to be that (1) for the purpose of describing a population, the mean, variance, median, 10th and 90th percentiles using the mapped utilities should closely approximate that of using the observed utilities, and (2) for the purpose of estimating group differences and association, the difference in mean utility between groups $(\mu_k - \mu_{k'})$ using the mapped utilities should closely approximate that of using the observed utilities. For completeness, we would also report measures of individual-level prediction errors such as mean squared error (MSE), mean absolute error (MAE) and intraclass correlation coefficient (ICC). We used the ANOVA estimator of ICC.(24) In this context, there are three sets of pairwise comparison because three mapping methods were considered. Each set of pairwise comparisons consisted of one observed and one mapped utility for each participant. We went on to perform a simulation and split-sample validation.

Simulation

The purpose of the simulation was to compare (1) the distribution parameters and (2) the mean utilities in strata, defined by Z, and the regression coefficient estimate of health utility Y on Z ($(\hat{\beta}_{Y|Z})$) on assumption of linear trend. Individual-level prediction errors were also reported. The population consists of three strata of equal size, coded as Z = 1, 2 or 3, with Z = 3 being the stratum with the best health and Z = 1 being the stratum with the worst health. The

predictor scores are generated by setting $X = 20 + 5 \times Z - Poisson(\lambda = 5)$, which resembled the distribution of the Physical Health domain score of the WHOQOL-BREF, with a ceiling at 35. In the simulated data to be shown in Figure 1, the values ranged from 8 to 35 (i.e. 28 levels). Health utility Y, was generated by a normal distribution, with mean equal to either $0.45 + 0.2Z$ (setting (i)) or $0.55 + 0.2Z$ (setting (ii)), standard deviation equal to 0.1, and values > 1 were recoded to 1. Under simulation settings (i) and (ii), 25% and 41% of observations had health utility above 1 before recoding, respectively. After recoding, the correlation between X and Y was 0.74 and 0.70 and that between Z and Y was 0.85 and 0.79 under simulation settings (i) and (ii), respectively. The skewness in Y was -0.5 and -0.9 under settings (i) and (ii), respectively. To assess the impact of ties and number of levels in the predictor scale, we further generated a coarsened version of the predictor (X-coarse), by recoding the x values to 10 levels only (≤17=1, 18-19=2, 20-21=3,…, 34-35=10). We ran 1000 replicates of the simulation, each with a sample size of 500. Within each replication, the mapping methods were applied to the new data to generate new mapping results.

Split-sample Validation

Participants were randomly assigned to the training and validation samples. To evaluate the performance of the mapping methods in describing the population, we tabulated and compared the mean, SD, percentiles, and variance ratio, $Var(\hat{Y})/Var(Y)$, where $\hat{Y}$ is the predicted health utility, of the observed EQ-5D-5L utilities and the three sets of mapped utilities. In addition, we arbitrarily

defined utility ≤ 0.8 as low utility and normal otherwise. This value of 0.8 was approximately the 25th percentile in the present study. The proportion of participants with low utility based on the observed utilities and three sets of mapped utilities were compared.

To evaluate the performance of the mapping methods in estimating group differences and association, we performed several least squares regression analyses. The first model has the observed utilities as dependent variable, and demographic variables (age and gender) as the independent variables. The second model has observed utilities as the dependent variable and the 8 domain scores of SF-36v2 as the independent variables. The third model has observed utilities as the dependent variable and self-reported morbidity on a disease inventory as the independent variable. We then repeated the above by replacing the observed utilities with one of three sets of mapped utilities, giving a total of 12 regression analyses. The regression coefficients and the acceptance/rejection of null hypotheses based on the observed and mapped utilities were compared. Individual-level prediction errors were also reported.

Results

Simulation

Figure 1 illustrates the cumulative distribution of observed and MRM-mapped health utilities under simulation setting (i) in a population of one million people generated according to the parameters described in the Methods section. The CDF of the mapped utilities follows that of the observed health utilities closely. A similar CDF pattern was observed for simulation setting (ii) (details not shown).

Evaluation of Mapping Methods – Describing the Population

Table 1 shows the mean of parameters over 1000 simulation runs under two settings, each of which had a sample size n=500. All mapping methods estimated the observed mean health utility accurately in both settings. However, the MRM outperform OLS method and EPM in several aspects. For instance, the MRM obtained a variance and percentiles close to the observed health utilities. The ratios of variance of MRM-, OLS method- and EPM-mapped utilities against the observed utilities were 0.995, 0.525 and 0.831, respectively, in setting (i) and 0.994, 0.427 and 0.865, respectively, in setting (ii) . The MRM underestimated the variance by less than 1% in the two settings. In addition, the MRM outperformed the OLS method and EPM in terms of accuracy in estimating the percentiles of the observed distribution. Even when the predictor data were recoded to a coarser (10-point) scale, the MRM still outperformed the OLS method in both settings and outperformed the

EPM in setting (i) in terms of estimating the true variance and the 10th, 50th and 90th percentiles.

Evaluation of Mapping Methods – Describing the Association

In terms of using mapped data to estimate association with covariate Z, the MRM provided regression coefficients closer to the true regression coefficients than the other mapping methods. The MRM also more accurately estimated the mean utility values in all three strata in both simulation settings than the other two methods. In both simulation settings, the OLS method showed the expected pattern of bias in estimating the mean utility by health strata. For example, in setting (i), the OLS method overestimated the mean utility for a poor health state (Z=1, 0.699 versus 0.650) and underestimated the mean utility for a good health state (Z=3, 0.947 versus 0.980; Table 1). Even when the predictor data were recoded to a coarser (10-point) scale, most of the above observations still held.

Individual-Level Prediction Error

Where individual-level prediction error was concerned, the OLS method had the smallest mean squared errors (Table 1). On the other hand, MRM had the smallest mean absolute errors.

Split-Sample Validation

Participants and Missing Data

There was a total of 1343 participants, with 913 participants from the general population, 223 from NHCS and 207 from NUH. The overall response rate in the general population survey was 21% and this is not unexpected for door-to-door surveys. We did not keep track of the response rate in the clinic samples. However, in a similar study assessing health-related quality of life in the same diabetes clinic, response rate was 32%.(25) Out of 1343 participants, 1318 (98.1%) participants had no more than one missing value per WHOQOL-BREF domain and 3 (0.2%) participants had missing value on other variables such as the SF-36v2. The item non-responses of these 1315 participants were imputed using the mean substitution method described earlier. Eighty four (6.4%) of the 1318 participants had one missing value within the Physical Health domain of the WHOQOL-BREF, mostly in item 18 about "satisfied with your capacity for work" (n=73); people who were not employed tended not to answer this question. The present manuscript analysed the 1315 participants, with 658 participants in the training and 657 in the validation samples. The participants comprised 755 men (57%), 616 Chinese (47%), 539 Malay (41%), 154 Indian (12%) and 6 of other race (0.5%) with mean age 52 (range: 24 to 90) years. Participants completed the survey in English (454, 35%), Chinese (442, 34%) or Malay languages (419, 32%). The percentage of participants with primary, secondary or tertiary education was 35%, 38% and 27%, respectively.

Descriptive and Exploratory Data Analyses in Training Dataset

Distribution of WHOQOL-BREF, EQ-5D-5L and SF-36v2 scores in the training sample are given in Table 2. In the training sample, the Physical Health, Psychological, Social Relationships, and Environment domain scores are integers in the ranges of 10 to 35, 9 to 30, 5 to 15, and 11 to 40, respectively. The observed EQ-5D-5L utility and WHOQOL-BREF domains scores covered a broad range. However, the lowest possible EQ-5D-5L utility and the WHOQOL-BREF domain scores were not observed.

A non-linear relation between WHOQOL-BREF Physical Health domain and EQ-5D-5L utility was found: as the Physical Health domain score increased, the increase in utility score levelled off. The non-linearity can be captured by a 2-degree fractional polynomial with power terms -2 and 0.5 (each $p<0.001$). Having included the Physical Health domain scores, none of the other three domain scores were associated with utility (each $p>0.40$). The model with all four domain scores (2 degrees of freedom (DF) for Physical Health and 1 DF for each of the other domains) and the model with only the Physical Health domain (2 DF) gave adjusted R-squared of 0.449 and 0.451, respectively. Hence, we went with the model that included only the Physical Health domain.

Estimation of Predicted Utilities in Training Dataset

The mapped utilities according to the MRM, OLS method and EPM are shown in Table 3 for the range of Physical Health domain scores (10 to 35) observed in this study. The OLS mapping equation is:

$$\text{EQ-5D utility} = 0.2621 - 46.8768 \times \text{Physical}^{-2} + 0.1327 \times \text{Physical}^{0.5}$$

Despite the non-linear functional form, if the Physical Health score is at its maximum score of 35, the OLS-mapped utility score at being 1.009 exceeded the full health utility. This was constrained to a maximum value of 1.0 to conform with the full health utility.

The OLS-mapped utilities tend to be higher than the MRM-mapped utilities, except when the MRM-mapped utilities equalled or was near the full health utility of 1 at Physical Health domain score > 28. Except in the lower end of Physical Health domain scores (10 to 13), the MRM-mapped utilities tended to be slightly higher than the EPM-mapped utilities.

Evaluation of Mapping Methods – Describing the Population

Table 4 shows the mean, SD, percentiles of the observed and three series of mapped utilities in the validation sample. The OLS method underestimated the observed SD (0.095 versus 0.145). The EPM slightly underestimated the observed mean utilities (0.843 versus 0.885). The MRM performed well in both

regards. Its variance ratio against the observed variance was 0.959. The MRM also gave percentiles that more closely followed the observed percentiles than the other two methods. According to the observed utility, 24.7% (95% CI: 21.4% to 28.1%) of the participants had a low (≤0.8) level of utility. Using the MRM, OLS method and EPM, 28.8% (25.3% to 32.4%), 14.0% (11.4% to 16.9%) and 44.3% (40.5% to 48.2%) had a low level of utility, respectively.

Evaluation of Mapping Methods – Describing the Association with Age and Gender

In assessing the association with age and gender, the observed utilities were negatively associated with age per five-year band (regression coefficient -15 x 10^{-3}; p<0.001) but not with gender (6.9 x 10^{-3}; p=0.53, Table 5). All three mapping methods gave similar findings.

Evaluation of mapping methods – Describing the Association with SF-36v2 Scale Scores

The observed EQ-5D-5L utility score had strong association with the PF and BP scores of the SF-36v2 (each p<0.001), was somewhat associated with GH (p=0.015) and MH (p=0.027), and clearly had no association with SF (Table 6). All three mapping methods similarly captured the strong association with PF

and lack of association with SF. However, all of them somewhat underestimated the association with BP and overestimated the association with VT. The OLS method gave standard errors that were smaller (range: 1.3 (BP) to 2.0 x 10^{-4} (MH), not counting the intercept) than those from the observed utility (2.3 (BP) to 3.5 (MH) x 10^{-4}) and the other two mapping methods.

Evaluation of Mapping Methods – Describing the Association with Self-reported Morbidity

Among conditions on an inventory, 6 conditions had prevalence above 5% (Table 7). Co-morbidities among patients with diabetes mellitus were common. Having controlled for co-morbidity, diabetes mellitus was not associated with the observed utility score (1.5 x 10^{-3}; p=0.92, Table 7). Asthmatic patients had somewhat reduced utility but that was statistically non-significant (-38 x 10^{-3}; p=0.096). It should be noted that, although the association between observed utility score and diabetes mellitus were not statistically significant, the association between mapped utility scores and diabetes mellitus were statistically significant across all three sets of mapped utility scores (each $p<0.01$). Furthermore, the OLS method showed a reduced utility level in patients with asthma that is close to conventional level of statistical significance (-26x10^{-3}; p=0.052). Other results were similar across methods.

Individual-Level Prediction Error

The OLS method had the lowest mean squared error while the MRM had the lowest mean absolute error. The EPM had larger prediction errors on both accounts. The ICCs between the observed utilities and the MRM-, OLS-method, and EPM-mapped utilities were 0.60 (95% CI: 0.55 to 0.65), 0.58 (0.53 to 0.63), and 0.56 (0.51 to 0.62), respectively. This suggests fair/good agreement at the individual-level.

Discussion

This is the first manuscript to map the WHOQOL-BREF to EQ-5D-5L, two widely used HRQoL measures. We have now enabled numerous studies that employ the WHOQOL-BREF alone without concurrently including a preference-based HRQoL measure to produce such a score that is critical for inclusion in cost-utility studies. In addition, we proposed a new mapping method, which we call the MRM. We demonstrated by simulation and split-sample validation that this new method performs better than the OLS method and the EPM.

In this article, we have focused our comparison on OLS method and EPM because the former is the most commonly used method so far and the latter is conceptually similar to our proposed method. Nonetheless, there are other alternative methods. For example, Gray et al.(26) used the multinomial regression approach to map SF-12 responses to the five dimensions of the EQ-5D, which were then used to generate the utility index, and Sullivan and Ghushchyan used Tobit regression to map the SF-12 scores to the EQ-5D

utility index.(27) In these two examples, the observed responses to EQ-5D dimensions or the observed EQ-5D utility index were considered the manifestation of latent variables. Despite more statistical sophistication, these regression methods did not appear to perform better than the OLS method in mapping to the utility index,(26, 27) even though the indirect mapping via the EQ-5D dimensions did provide richer descriptive information. As regression methods other than OLS method have not provided much improvement in mapping, we consider the alternative approach along the line of EPM.

All three mapping methods estimated the mean and median reasonably accurately. However, in terms of describing the lower end of the distribution and the proportion with low EQ-5D-5L utility, the MRM out-performed the other two methods in both simulation and split-sample validation. It is well known that the variability of utility scores generated by OLS mapping is underestimated. But its implication on CUA appears to have received limited attention. Since OLS mapping replaces high and low values with the corresponding conditional means, mean utilities for the group with a good (bad) health state tends to be underestimated (overestimated). As a result, the gradient of utilities across ordered health states tends to be underestimated, which leads to biased estimation of the gain in QALY and incremental cost-utility ratio. This pattern was previously reported in, for example, cancer patients classified according to performance status (28) and migraine patients classified according to number of headache days per month.(29) In both simulation settings, we found that the MRM and EPM were successful in preserving the differences in mean utilities

between groups and the regression coefficient. In contrast, the OLS method underestimated the differences, with the regression coefficient being under-estimated by 25% to 28%.

In terms of estimating association and the mean difference between groups, the relative accuracy of the MRM was more visible in simulation than in the split-sample validation. In the split-sample validation, the three methods captured most of the major findings using the observed EQ-5D-5L utility. It was true that the OLS method gave smaller variability, but it did not differ vastly from the other methods in terms of rejecting hypotheses. We do note that there was a discrepancy in the covariate-adjusted association between observed EQ-5D-5L utility score and diabetes mellitus and the covariate-adjusted association between mapped EQ-5D-5L utility score and diabetes mellitus. We believe that this discrepancy is not due to an underestimation of the variance produced by the MRM as we have demonstrated that the underestimation with the MRM using WHOQOL as predictor is negligible. It should be noted that the multiple regression analysis (Table 7) adjusted for comorbidities. As such, the non-significant regression coefficient for diabetes mellitus in the analysis of observed EQ-5D-5L utility scores represented the non-significant difference between healthy people and persons with diabetes who had no comorbidity, which was only a small fraction of persons with diabetes mellitus in our study population. Ceiling effects are pronounced with the observed EQ-5D-5L utility scores. Pooling the training and validation datasets, 16 of 25 persons (64%)

with diabetes mellitus but no comorbidity were at the EQ-5D-5L ceiling. In contrast, none of this group of participants reached the WHOQOL-Physical Health score ceiling. Since ceiling effect suppresses differences in people who are relatively well, the mapped utility scores were more likely to show a significant coefficient for diabetes mellitus in comorbidity-adjusted analysis. Similarly, despite statistical significance in all four sets of analyses, the absolute values of the regression coefficients on the General Health score of the SF-36 (Table 6) and heart and renal diseases (Table 7) in the analyses of observed EQ-5D-5L utilities were smaller than those in the analyses of mapped utilities, by roughly two to three standard errors. That the WHOQOL-BREF scores and mapped EQ-5D-5L utilities had less ceiling effect than the observed EQ-5D-5L utilities may have also contributed to these differences. The present study did not include an assessment of the impact of using mapped versus actual utilities on the incremental cost-utility ratio because that will be similar to looking at group differences (i.e. differences in utilities between patients on new and old treatments).

With regards to individual-level prediction error, as expected, both the simulation and validation analysis showed that the mean squared error was lowest in the OLS-based mapping. However, both the OLS method and EPM had higher mean absolute error than the MRM. In terms of intraclass correlation, the three methods performed quite similarly, with the MRM showing slightly stronger agreement with the observed data than the other two methods.

There was no consistent pattern to clearly indicate which method is more accurate in making individual-level predictions.

Kernel smoothing is no trivial task. It involves decisions that are somewhat subjective, such as choice of kernel function, band width, and handling of boundary effects. That the EPM performed somewhat less well than the MRM in this study does not necessarily mean that the methodological framework is poor. The issue with the EPM is that different analysts may produce somewhat different results depending on the choice of the parameters described above. This leads us to point out that the MRM is simple to use and does not involve any subjective modelling choices. Most importantly, all analysts will produce the same mapping result given the same dataset. Simplicity and reproducibility are thus two relative advantages of the MRM.

The accuracy of the MRM is reduced as the number of levels on the predictor scale decreases. The observed number of predictor levels in this study was 26 (integers from 10 to 35), although the maximum possible number for the scale was 39 (integers from 7 to 35). In the case study and in simulation settings with similar data patterns, the MRM performed well. We note that profile-based HRQoL measures that are used to map to utilities often have more levels than we evaluated. For example, the observed number of levels in the Functional Assessment of Cancer Therapy (FACT) – General and the FACT-Breast in a mapping study were 74 and 96, respectively.(30) As such, we expect that the

proposed mapping method will have wide applicability, even though caution is needed in scales with fewer levels than ours.

As proposed by Chan et al.,(31) an adjustment factor can be derived to inflate the variance of mapped utilities. However, in both simulation and split-sample validation, we have found that for use of the WHOQOL-BREF (or other scales with similar number of levels of measurement) as predictor of health utility, the ratio of variance was very close to one, allowing us to practically ignore such an adjustment. However, as the predictor scale becomes coarser, the adjustment should be considered.

We acknowledge that this study has a few limitations. First, the mapping does not cover the full theoretical raw score range (7 to 35) of WHOQOL-BREF Physical Health domain and is restricted only to the observed raw score range of 10 to 35. This may limit the applicability of the mapping in samples with very poor WHOQOL-BREF Physical Health scores. Despite our effort to include a combination of general population and clinic participants, we were unable to capture individuals with extremely poor HRQoL. It is expected that individuals with such poor physical health may be found either in inpatient or nursing home settings and will inevitably be missed in any study conducted in the community or ambulatory care setting. Second, there may be some concern that the mapped EQ-5D-5L scores reflect only the Physical Health and not the other three domains of WHOQOL-BREF. However, we have found that including all three domains did not improve the model fit of the regression between

WHOQOL-BREF and EQ-5D-5L utilities. Hence, our modelling choice was empirically supported. In a previous study, we have also found that only four of five subscales of the Functional Assessment of Cancer Therapy - Breast (FACT-B) were associated with EQ-5D-5L utilities and the mapping algorithm was based on these four subscales.(30) Third, there may be some concern over the generalizability of the findings from Singapore to other countries. However, the design of WHOQOL-BREF has intentionally taken cross cultural differences into account, including only those items that are applicable across various countries. As such, we believe that generalizability of our findings to other countries is unlikely to be an issue. Fourth, this study has covered a general population and two outpatient populations (heart disease and diabetes). Further assessment of the performance of the mapping algorithm in other clinical conditions and inpatient/ non-ambulatory patient situations is recommended.

Conclusion

We have provided the first algorithm for translating profile-based WHOQOL-BREF Physical Health domain scores into utilities that can be incorporated into cost-utility analyses for the evaluation of new drugs or other healthcare interventions. The proposed MRM has practical and numerical advantages although all three mapping methods were approximately valid in generating EQ-5D-5L utilities in this Singaporean study. In scenarios similar to the mapping of WHOQOL-BREF to the EQ-5D-5L, the MRM outperformed OLS

method and EPM. The simplicity and reproducibility of the MRM makes it an attractive alternative to current methods. To access de-identified data used in this analysis, please email the corresponding author for a data analysis proposal form, which is to be duly completed and returned to the corresponding author.

REFERENCES

1. Drummond MF, Sculpher MJ, Claxton K, et al. Methods for the Economic Evaluation of Health Care Programmes. Oxford University Press; 2015
2. Herdman M, Gudex C, Lloyd A, 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
3. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of health economics 2002;21:271-292
4. Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health and quality of life outcomes 2003;1:54
5. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. UK: National Institute for Health and Care Excellence; 2013
6. Fayers PM, Hays RD. Should linking replace regression when mapping from profile-based measures to preference-based measures? Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2014;17:261-265

7. Young TA, Mukuria C, Rowen D, et al. Mapping Functions in Health-Related Quality of Life: Mapping from Two Cancer-Specific Health-Related Quality-of-Life Instruments to EQ-5D-3L. Medical Decision Making 2015;35:912-926
8. Versteegh MM, Leunis A, Luime JJ, et al. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. Medical Decision Making 2012;32:554-568
9. Brazier JE, Yang Y, Tsuchiya A, et al. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European journal of health economics : HEPAC : health economics in prevention and care 2010;11:215-225
10. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998;28:551-558
11. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med 1995;41:1403-1409
12. Dorans NJ. Linking scores from multiple health outcome instruments. Qual Life Res 2007;16 Suppl 1:85-94
13. Holland PW, Thayer DT. The Kernel Method of Equating Score Distributions. Educational Testing Service.; 1989
14. Huang IC, Frangakis C, Atkinson MJ, et al. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. Health services research 2008;43:327-339
15. Petrou S, Rivero-Arias O, Dakin H, et al. Preferred Reporting Items for Studies Mapping onto Preference-Based Outcome Measures: The MAPS Statement. Medical Decision Making 2015;35:NP1-NP8
16. World Health Organization Division of Mental Health. WHOQOL-BREF : introduction, administration, scoring and generic version of the assessment : field trial version. Geneva: World Health Organization; 1996
17. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2012;15:708-715
18. Tsuchiya A, Ikeda S, Ikegami N, et al. Estimating an EQ-5D population value set: the case of Japan. Health economics 2002;11:341-353

19. Sow WT, Wee HL, Wu Y, et al. Normative Data for the Singapore English and Chinese SF-36 Version 2 Health Survey. Ann Acad Med Singapore 2014;43:15-23
20. Van Kerm P. Kernel-smoothed cumulative distribution function estimation with akdensity. Stata Journal 2012;12:543-548
21. Cowling A, Hall P. On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation. Journal of the Royal Statistical Society Series B (Methodological) 1996;58:551-563
22. Hall P, Wehrly TE. A Geometrical Method for Removing Edge Effects from Kernel-Type Nonparametric Regression Estimators. Journal of the American Statistical Association 1991;86:665-672
23. Silverman BW. Density estimation for statistics and data analysis. CRC press; 1986
24. Donner A. A Review of Inference Procedures for the Intraclass Correlation Coefficient in the One-Way Random Effects Model. International Statistical Review 1986;54:67-82
25. Venkataraman K, Tan LS, Bautista DC, et al. Psychometric Properties of the Problem Areas in Diabetes (PAID) Instrument in Singapore. PloS one 2015;10:e0136759
26. Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Medical decision making : an international journal of the Society for Medical Decision Making 2006;26:18-29
27. Sullivan PW, Ghushchyan V. Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. Medical decision making : an international journal of the Society for Medical Decision Making 2006;26:401-409
28. Cheung YB, Thumboo J, Gao F, et al. Mapping the English and Chinese versions of the Functional Assessment of Cancer Therapy-General to the EQ-5D utility index. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2009;12:371-376
29. Gillard PJ, Devine B, Varon SF, et al. Mapping from disease-specific measures to health-state utility values in individuals with migraine. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2012;15:485-494

30. Cheung YB, Luo N, Ng R, et al. 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 2014;12:180
31. Chan KK, Willan AR, Gupta M, et al. Underestimation of uncertainties in health utilities derived from mapping algorithms involving health-related quality-of-life measures: statistical explanations and potential remedies. Medical decision making : an international journal of the Society for Medical Decision Making 2014;34:863-872

Table 1. Mean of parameter estimates in 1000 simulation runs (n=500 each) under two settings.

Parameters	Observed Values	Mapped Values			
		Mean Rank [a]	Mean Rank (Coarsened) [b]	Regression (OLS)	Equipercentile
Setting (i)					
Mean	0.826	0.826	0.825	0.824	0.803
Variance	0.025	0.025	0.025	0.013	0.021
P10	0.556	0.599	0.586	0.672	0.605
P50	0.850	0.849	0.836	0.827	0.813
P90	1.000	1.000	1.000	0.983	0.983
Mean | Z=1	0.650	0.646	0.648	0.699	0.641
Mean | Z=2	0.847	0.849	0.848	0.826	0.811
Mean | Z=3	0.980	0.983	0.981	0.947	0.955
$(\hat{\beta}_{Y\|Z})$	0.165	0.168	0.167	0.124	0.157
Mean squared error	-	0.0128	0.0129	0.0111	0.0120
Mean absolute error	-	0.0810	0.0815	0.0815	0.0825
Variance ratio	-	0.995	0.980	0.525	0.831
Setting (ii)					
Mean	0.892	0.892	0.892	0.889	0.865
Variance	0.016	0.016	0.016	0.007	0.014
P10	0.696	0.699	0.686	0.777	0.698
P50	0.950	0.949	0.935	0.893	0.889
P90	1.000	1.000	1.000	1.000	0.990
Mean | Z=1	0.750	0.746	0.748	0.797	0.730
Mean | Z=2	0.930	0.934	0.933	0.892	0.886
Mean | Z=3	0.997	0.997	0.997	0.976	0.980
$(\hat{\beta}_{Y\|Z})$	0.124	0.126	0.124	0.089	0.125
Mean squared error	-	0.0098	0.0097	0.0081	0.0094
Mean absolute error	-	0.0640	0.0642	0.0660	0.0690
Variance ratio	-	0.994	0.976	0.427	0.865

[a] Mapped values using Mean Rank mapping of raw predictor scores

[b] Mapped values using Mean Rank mapping after recoding predictor scores to only 10 levels

Table 2. Distribution of WHOQOL-BREF, EQ-5D-5L and SF-36v2 scores in the training sample (n=658).

Measure	**Mean (SD)**	**Observed Range**
WHOQOL-BREF Physical Health	27.5 (4.55)	10 to 35
WHOQOL-BREF Psychological	22.9 (3.46)	9 to 30
WHOQOL-BREF Social Relationships	11.7 (1.79)	5 to 15
WHOQOL-BREF Environment	30.4 (4.39)	11 to 40
EQ-5D-5L utility	0.9 (0.15)	-0.02 to 1
SF-36v2 Physical Functioning	85.2 (21.13)	0 to 100
SF-36v2 Role Physical	82.4 (24.15)	0 to 100
SF-36v2 Bodily Pain	80.3 (21.43)	0 to 100
SF-36v2 General Health	67.7 (20.05)	0 to 100
SF-36v2 Vitality	66.1 (19.08)	0 to 100
SF-36v2 Social Functioning	84.9 (22.08)	0 to 100
SF-36v2 Role-Emotional	86.9 (21.32)	0 to 100
SF-36v2 Mental Health	78.3 (17.69)	5 to 100

Table 3. Mapped EQ-5D-5L Utilities to WHOQOL-BREF Physical Health Domain Using Mean Rank-, Regression- and Equipercentile Approaches.

WHOQOL-BREF Physical Health Domain	**Mean rank method**	**Regression (OLS) method**	**Equipercentile method**
10	0.05	0.21	0.20
11	0.23	0.31	0.25
12	0.29	0.31	0.30
13	0.34	0.46	0.35
14	0.43	0.52	0.41
15	0.48	0.57	0.47
16	0.56	0.61	0.52
17	0.60	0.65	0.57
18	0.63	0.68	0.60
19	0.66	0.71	0.63
20	0.68	0.74	0.66
21	0.70	0.76	0.68
22	0.72	0.79	0.69
23	0.73	0.81	0.71
24	0.74	0.83	0.72
25	0.78	0.85	0.74
26	0.81	0.87	0.77
27	0.82	0.89	0.80
28	0.97	0.90	0.82
29	1.00	0.92	0.86

WHOQOL-BREF Physical Health Domain	**Mean rank method**	**Regression (OLS) method**	**Equipercentile method**
30	1.00	0.94	0.98
31	1.00	0.95	0.99
32	1.00	0.97	0.99
33	1.00	0.98	1.00
34	1.00	1.00	1.00
35	1.00	1.00	1.00

Rounded off to two significant figures.

Table 4. Mean, Standard Deviation (SD), Percentiles, Mean Squared Errors (MSE), Mean Absolute Errors (MAE) and Intraclass Correlation Coefficient (ICC) of the Observed and Mean Rank-, Regression- and Equipercentile-mapped EQ-5D-5L Utilities in the Validation Sample (n=657).

Mapping Method	**EQ-5D-5L Utilities**								**Individual Level Prediction Errors**		**Agreement**
	Mean (SD)	5^{th}	10^{th}	25^{th}	Median	75^{th}	90^{th}	95^{th}	MSE	MAE	ICC
Observed	0.885 (0.145)	0.649	0.700	0.813	1.000	1.000	1.000	1.000	-	-	-
Mean Rank	0.883 (0.142)	0.657	0.697	0.777	0.973	1.000	1.000	1.000	0.016	0.084	0.60
Regression (OLS)	0.886 (0.095)	0.711	0.764	0.851	0.905	0.952	0.981	0.995	0.013	0.091	0.58
Equipercentile	0.843 (0.140)	0.630	0.676	0.744	0.822	0.988	0.996	0.998	0.018	0.095	0.56

Rounded off to three decimal places.

Table 5. Association between Observed, Mean Rank-, Regression- and Equipercentile- mapped EQ-5D-5L Utilities with Age (5-year band) and Gender (n=657).

EQ-5D-5L Utilities	**Beta coefficient ($x10^{-3}$)**	**Standard error ($x10^{-3}$)**	**p**	**95% Confidence Interval ($x10^{-3}$)**
Observed				
Age	-15	1.7	<0.001	-18 to -11
Gender	6.9	11	0.53	-15 to 2.8
Constant	980	14	<0.001	950 to 1000
Mean Rank-mapped				
Age	-17	1.7	<0.001	-20 to -14
Gender	-14	10	0.19	-35 to 7.0
Constant	100	13	<0.001	980 to 1030
Regression-mapped				
Age	-10	1.1	<0.001	-13 to -8.3
Gender	-8.1	7.1	0.25	-22 to 5.8
Constant	960	8.7	<0.001	940 to 980
Equipercentile-mapped				
Age	-17	1.6	<0.001	-20 to -14
Gender	-3.3	10	0.75	-24 to 17
Constant	960	13	<0.001	930 to 980

Rounded off to two significant figures.

Table 6. Association between Observed, Mean Rank-, Regression- and Equipercentile- mapped EQ-5D-5L Utilities with SF-36v2 Scale Scores (n=647).

EQ-5D-5L Utilities	**Beta coefficient ($x10^{-4}$)**	**Standard error ($x10^{-4}$)**	**p**	**95% Confidence Interval ($x10^{-4}$)**
Observed				
PF	18	2.8	<0.001	12 to 23
RP	4.1	2.7	0.14	-1.3 to 9.4
BP	19	2.3	<0.001	14 to 24
GH	7.2	3.0	0.015	1.4 to 13
VT	3.9	3.3	0.24	-2.5 to 10
SF	-0.7	2.5	0.79	-5.6 to 4.3
RE	4.3	2.8	0.12	-1.1 to 9.7
MH	7.8	3.5	0.027	0.9 to 15
Constant	380	22	<0.001	34 to 42
Mean Rank Method				
PF	16	2.4	<0.001	11 to 20
RP	7.0	2.3	0.0040	23 to 12
BP	3.5	2.0	0.094	-0.60 to 7.5
GH	19	2.6	<0.001	14 to 24
VT	15	2.8	<0.001	8.8 to 20
SF	0.5	2.2	0.82	-3.9 to 4.9
RE	0.6	2.4	0.014	1.2 to 11
MH	1.0	3.0	0.74	-5.1 to 7.1
Constant	380	19	<0.001	340 to 420
Regression (OLS) Method				
PF	11	1.6	<0.001	7.9 to 14
RP	4.0	1.5	0.0090	1.0 to 7.0
BP	3.3	1.3	0.012	7.0 to 6.0

EQ-5D-5L Utilities	**Beta coefficient ($x10^{-4}$)**	**Standard error ($x10^{-4}$)**	**p**	**95% Confidence Interval ($x10^{-4}$)**
GH	11	1.7	<0.001	7.8 to 14
VT	9.2	1.8	<0.001	5.6 to 13
SF	1.2	1.4	0.39	-1.6 to 40
RE	2.6	1.6	0.099	-0.5 to 5.6
MH	4.4	2.0	0.026	0.5 to 8.3
Constant	530	12	<0.001	51 to 55
Equipercentile Method				
PF	13	2.3	<0.001	8.0 to 17
RP	7.1	2.3	0.0020	2.7 to 12
BP	3.7	2.0	0.062	-0.2 to 7.6
GH	20	2.5	<0.001	16 to 25
VT	15	2.7	<0.001	9.1 to 20
SF	0.20	2.1	0.92	-4.0 to 4.4
RE	4.7	2.3	0.045	0.1 to 9.2
MH	3.7	3.0	0.21	-2.1 to 9.5
Constant	340	19	<0.001	310 to 380
	PF – Physical Functioning, RP – Role-Physical, BP – Bodily Pain, GH – General Health, VT – Vitality, SF – Social Functioning, RE – Role-Emotional, MH – Mental Health Rounded off to two significant figures.			

Table 7. Association between Observed, Mean Rank-, Regression- and Equipercentile- mapped EQ-5D-5L Utilities with Self-reported Morbidity (n=654).

EQ-5D-5L Utilities	**Beta coefficient ($x10^{-3}$)**	**Standard error ($x10^{-3}$)**	**p**	**95% Confidence Interval ($x10^{-3}$)**
Observed				
Diabetes mellitus	1.5	15	0.92	-29 to 32
Hypertension	-47	14	0.0010	-75 to -20
Heart disease	-39	14	0.0060	-67 to -11
Asthma	-38	22	0.096	-82 to 6.6
Mental illness	-83	16	<0.001	-110 to -52
Renal disease	-55	23	0.016	-100 to -10
Constant	930	6.8	<0.001	920 to 940
Mean Rank Method				
Diabetes mellitus	-47	14	0.0010	-74 to -20
Hypertension	-44	12	<0.001	-68 to -20
Heart disease	-79	12	<0.001	-100 to -55
Asthma	-26	19	0.19	-65 to 13
Mental illness	-59	14	<0.001	-86 to -32
Renal disease	-92	20	<0.001	-130 to -53
Constant	950	5.8	<0.001	940 to 0.96
Regression (OLS) Method				
Diabetes mellitus	-27	9.3	0.0040	-45 to -8.3
Hypertension	-36	8.3	<0.001	-52 to -20
Heart disease	-43	8.4	<0.001	-60 to -27
Asthma	-26	13	0.052	-52 to 0.2
Mental illness	-38	9.2	<0.001	-56 to -20
Renal disease	-64	13	<0.001	-91 to -38
Constant	930	4.0	<0.001	920 to 940

EQ-5D-5L Utilities	**Beta coefficient ($x10^{-3}$)**	**Standard error ($x10^{-3}$)**	**p**	**95% Confidence Interval ($x10^{-3}$)**
Equipercentile Method				
Diabetes mellitus	-41	14	0.0020	-68 to -15
Hypertension	-55	12	<0.001	-79 to -31
Heart disease	-68	12	<0.001	-92 to -44
Asthma	-30	19	0.12	-68 to 7.8
Mental illness	-58	13	<0.001	-85 to -32
Renal disease	-83	20	<0.001	-120 to -45
Constant	910	5.8	<0.001	900 to 920

Rounded off to two significant figures.

Online Appendix 1. Numerical features of the mean rank mapping method

Let $x_{(1)} < x_{(2)} < \cdots < x_{(K-1)} < x_{(K)}$ be the K distinct values of x, n_k be the number of observations with value $x_{(k)}$, and $N = \sum n_k$ is the study sample size.

Define $y_i = \hat{y}_i + e_i$, where $\hat{y}_i$ is the mapped value and e_i is the mapping error for the i^{th} row of the sorted (x, y) data as described in the manuscript.

Define $R_k = \sum_{j=1}^{k} n_j$ for $k \geq 1$, $R_0 = 0$, and $u_k = \sum_{i=R_{k-1}+1}^{R_k} y_i / n_k = \hat{y}_i$ for $R_{k-1} < i \leq R_k$ is the mean of the observed y's to be mapped to $x_{(k)}$ as described in the manuscript.

The mean of mapped values is identical to the mean of the observed y values:

$$\frac{1}{N}\sum_{i=1}^{N} \hat{y}_i = \frac{1}{N}\sum_{k=1}^{K} n_k u_k = \bar{y}$$

The variance of observed y values is:

$$Var(y_i) = \frac{\sum_{i=1}^{N} y_i^2}{N} - \bar{y}^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i + e_i)^2}{N} - \bar{y}^2 = \frac{\sum_{i=1}^{N} \hat{y}_i^2 + 2\sum_{i=1}^{N} \hat{y}_i e_i + \sum_{i=1}^{N} e_i^2}{N} - \bar{y}^2$$

$$= \left(\frac{\sum_{i=1}^{N} \hat{y}_i^2}{N} - \bar{y}^2\right) + \frac{2 \times 0}{N} + \frac{\sum_{i=1}^{N} e_i^2}{N} = Var(\hat{y}_i) + \frac{\sum_{i=1}^{N} e_i^2}{N}$$

As $\sum_{i=1}^{N} e_i^2 \geq 0$, $Var(y_i) \geq Var(\hat{y}_i)$.

The sum of squared mapping errors is:

$$\sum_{k=1}^{K} \sum_{i=R_{k-1}+1}^{R_k} (y_i - \hat{y}_i)^2$$

Let $y_{k,low}$ and $y_{k,up}$ be the y values on the $(R_{k-1} + 1)th$ and $(R_k)th$ row of the sorted data, respectively, and y_p $(y_{k,low} \leq y_p \leq y_{k,up})$ be the p^{th} percentile value defined as the y_i where $i = p \times N/100$ and $R_{k-1} < i \leq R_k$. The absolute mapping error for the p^{th} percentile is:

$$\left|y_p - u_k\right| \leq max\{u_k - y_{k,low}, y_{k,up} - u_k\} \leq (y_{k,up} - y_{k,low})$$

As the total number of levels K on the predictor variable X and the spread of the predictor values increase, the difference $(R_k - R_{k-1})$ and the width of the interval $(y_{k,up} - y_{k,low})$ decrease. So the

sum of squared mapping errors and the mapping errors for the percentiles also decrease. In the case of a unique x on the i^{th} row, $u_k = y_{k,low} = y_{k,up} = y_p$ and the mapped and observed p^{th} percentile exactly agree.