Does country-specific preference-weights matter in the choice of mapping algorithms? The case of mapping the Diabetes-39 onto eight country-specific EQ-5D-5L value sets

Concise title: <u>Mapping Diabetes-39 onto country-specific EQ-5D-5L tariffs</u>

Admassu N. Lamu[1], Gang Chen*[2], Thor Gamst-Klaussen[1], Jan Abel Olsen[1]

[1] Department of Community Medicine, University of Tromsø, N-9037 Tromsø, Norway
[2] Centre for Health Economics, Monash Business School, Monash University, VIC 3800, Australia

*Corresponding author: Room 295, 15 Innovation way, Clayton campus, Monash University, VIC 3800, Australia. Tel: +61-399050502, +61-425811029; Fax: +61-399058344; Email: gang.chen@monash.edu, chengang1029@gmail.com.

Does country-specific preference-weights matter in the choice of mapping algorithms? The case of mapping the Diabetes-39 onto eight country-specific EQ-5D-5L value sets

Abstract

Purpose: To develop mapping algorithms that transform Diabetes-39 (D-39) scores onto EQ-5D-5L utility values for each of eight recently published country-specific EQ-5D-5L value sets, and to compare mapping functions across the EQ-5D-5L value sets.

Methods: Data include 924 individuals with self-reported diabetes from six countries. The D-39 dimensions, age and gender were used as potential predictors for EQ-5D-5L utilities, which were scored using value sets from eight countries (England, Netherland, Spain, Canada, Uruguay, China, Japan and Korea). Ordinary least squares, generalized linear model, beta binomial regression, fractional regression, MM-estimation, and censored least absolute deviation were used to estimate the mapping algorithms. The optimal algorithm for each country-specific value set was primarily selected based normalized root mean square error (NRMSE), normalized mean absolute error (NMAE) and adjusted-r^2. Cross-validation with 5-fold approach was conducted to test the generalizability of each model.

Results: The fractional regression model with *loglog* as a link function consistently performed best in all country-specific value sets. For instance, the NRMSE (0.1282) and NMAE (0.0914) were the lowest, while adjusted-r^2 was the highest (52.5%) when the English value set was considered. Among D-39 dimensions, the *energy and mobility* was the only one that was consistently significant for all models.

Conclusions: The D-39 can be mapped onto the EQ-5D-5L utilities with good predictive accuracy. The fractional regression model, which is appropriate for handling bounded outcomes, outperformed other candidate methods in all country-specific value sets. However, the regression coefficients differed reflecting preference heterogeneity across countries.

Keywords: Mapping, Diabetes-39, EQ-5D-5L, HRQoL, Utility, QALY

Does country-specific preference-weights matter in the choice of mapping algorithms? The case of mapping the Diabetes-39 onto eight country-specific EQ-5D-5L value sets

Introduction

Condition-specific health-related quality of life (HRQoL) instruments are commonly regarded as being superior to generic HRQoL instruments at identifying changes in health [1]. However, for the purpose of comparing the relative cost-effectiveness of competing healthcare programmes across different disease areas, generic preference-based outcome measures are required. The most widely applied such instrument is the EQ-5D [2] that produces health state utilities for use in the calculation of quality-adjusted life years (QALYs). Given that clinical trials often do not collect data on patients' health state utilities, there is an increasing research interest in developing mapping algorithms that can estimate the relationship between a condition-specific measure of health, or 'source' instrument onto a generic preference-based HRQoL measure, or 'target' instrument [3; 4].

Diabetes is a common cause of premature death and disability, which imposes a large economic burden on the healthcare system [5]. In 2015, the International Diabetes Federation estimated about 415 million diabetic people (or 8.8% of adults aged 20-79) and 5 million deaths [6]. The rising prevalence of diabetes will lead to increased demand for healthcare services. Consequently, there is a growing need for evaluating the cost-effectiveness of many new diabetes interventions as compared to competing use of resource in other disease groups [4; 7].

The Diabetes-39 (D-39) is the most widely used non preference-based condition-specific instrument addressing the HRQoL of diabetic patients [8], while the EQ-5D is the most widely used generic preference-based instrument [9-11]. Over 17,000 studies using the EQ-5D instrument have been registered with the EuroQol Group by 2015 [11]. Further, the United Kingdom (UK) National Institute for Health and Care Excellence (NICE) [12] strictly requested that the EQ-5D should be used in all economic evaluations to ensure comparability between studies. Mapping studies aimed at generating algorithms for predicting EQ-5D utilities are also increasing, particularly after the NICE has endorsed mapping when the direct measure of EQ-5D utility is unavailable [13].

So far, only one study has developed mapping functions from D-39 to generic preference-based instruments including the five-level EQ-5D (EQ-5D-5L) [14]. A key limitation of this previous study is that it was based on an interim value set, which was a 'cross walk' between the earlier three-level

1

EQ-5D (EQ-5D-3L) and the EQ-5D-5L descriptive system [15]. Most recently, several country-specific EQ-5D-5L value sets have been published, including four Western countries (England, the Netherlands, Spain, Canada), three Asian countries (China, Japan, Korea) and one South American (Uruguay) [16-23]. Hence, the existing mapping algorithm for D-39 is becoming obsolete following the publication of these new EQ-5D-5L value sets.

This paper makes three important contributions to the literature. First, it produces optimal mapping functions for EQ-5D-5L value sets recently developed by direct elicitation of the general public. Second, we test several regression models. While the previous mapping algorithm for D-39 applied only ordinary least squares (OLS) and generalised linear model (GLM) [14], the current paper investigates into the merit of four other regression models. Third, we show that when preference based value sets differ across countries, the regression coefficients in the mapping functions will also differ. Thus, there is a need to derive country-specific mapping algorithms.

The present study aims to develop mapping algorithms that transform D-39 scores onto EQ-5D-5L utility values for each of eight recently published country-specific EQ-5D-5L value sets and to compare mapping functions across the EQ-5D-5L value sets. This study followed the recently developed checklist of minimum reporting requirement for mapping studies [24]: 'MApping Preference-based Measures reporting Standards (MAPS)'.

Methods
Data

Data were obtained from the Multi Instrument Comparison (MIC) study which is based on an online survey administered in six countries (Australia, Canada, Germany, Norway, UK, and United States of America) by a global panel company, CINT Australia Pty Ltd. The key aims of this large-scale study that includes more than 8,000 subjects were i) to compare health state utilities obtained from different generic preference-based instruments, and ii) to compare such values with values obtained from disease specific instruments. There was no missing value in the online survey. However, considering the lack of direct control in the online survey, several edit procedures (e.g. a comparison of duplicated questions, and removal of respondents whose recorded completion time shorter than 20 minutes) were conducted to ensure the quality of data. For further details on the MIC study, see Richardson et al. [25]. Among the seven chronic disease groups included in this comprehensive international study, the current paper is based on the diabetes group (N=924). The MIC data is an ideal source for developing mapping algorithms from condition-specific HRQoL onto generic preference-based values. So far it has been used to derive mapping functions in different chronic diseases, including asthma [26] depression [27], and heart diseases [28].

Measures of variables

The *EQ-5D* includes five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. The construction of the EQ-5D-5L retained the original dimensional structure of the EQ-5D-3L questionnaire and increased the descriptive systems to five levels of severity to address the problem of sensitivity and potential ceiling effects [29]. The five severity levels include no problems, slight problems, moderate problems, severe problems, and unable to/extreme problems. Here, we applied the new EQ-5D-5L value sets from eight countries [16-23]: England, the Netherlands, Spain, Canada, Uruguay, China, Japan, and Korea. For ease of exposition, the English value set would be emphasised, though results for other countries' value sets were discussed. The scale length differs across countries with the worst health state or the 'pits' (55555) ranges from -0.446 in the Netherlands to -0.025 in Japan [30].

The *D-39* includes 39-items, each with 7-response levels ranging from 1 (not affected at all) to 7 (extremely affected) [8]. The D-39 covers five domains: energy and mobility (15 items), diabetes control (12 items), anxiety and worry (4 items), social burden (5 items), and sexual functioning (3 items). Each item was reverse-coded, such that low values representing poor HRQoL. Domain scores were calculated by summing the item scores within each domain. Similarly, the total score of D-39 was obtained by summing the entire 39 item scores, where item scores are set equal to the rank order of the response. Finally, the total D-39 score and each subscale summary scores were linearly transformed onto a 0 to 1 scale: 0 indicating the worst, and 1 the best possible health state.[1]

It should be noted that the data set on which the modelling is based does not include subjects from any Asian countries. Thus, to the extent that people from Asia would describe their health problems differently along the instruments considered here, the regression coefficients in the mapping functions for the three Asian countries should be interpreted with somewhat more caution than those for the Western countries.

Statistical analyses and estimation

Descriptive analyses

Respondents' characteristics were described as mean (standard deviation, SD) unless otherwise indicated. The degree of conceptual overlap between the source (D-39) and target (EQ-5D-5L) instruments were examined using Spearman's correlation coefficients and a principal component analysis (PCA). PCA is a multivariate statistical technique used to reduce the number of variables in a data set into a smaller number of linearly uncorrelated 'principal components' that helps to

---

[1] For example, the transformed score for energy and mobility (EM) domain is calculated by subtracting the sum of reverse-coded responses across all 15 items (say, X) from its minimum score (15) and divided by the range (maximum (105) minus minimum scores). That is, the algorithm for computing the EM score is: (X-15)/(105-15).

investigate whether a number of items generate information about a more general underlying component [31]. The PCA was applied to explore and compare the underlying dimensional structure of the D-39 data and EQ-5D-5L information evident in these data. The number of principal components was decided according to eigenvalues, which indicate the amount of variance in the original variables explained by each principal component [32]. Only principal components with eigenvalues greater than 1 were considered in the exploratory PCA [33]. Rotation of the dimensions identified in the initial extraction is needed in order to obtain simple and interpretable components [34]. Rotation methods, which allows correlations between components, are referred to as oblique. Thus, to account for potential correlations among components, rotation was performed using an oblique Promax method [35].

Econometric techniques for mapping analysis
A direct mapping approach was adopted. The D-39 five domains, as well as age and gender, were the potential variables to predict EQ-5D-5L utilities. The final predictors were determined via stepwise forward selection that included only significant variables ($p < 0.05$). Predictors were also required to be logically consistent, meaning that poorer scores on a source instrument should lead to lower utility on the target instrument. Non-linearities were investigated, and they were included only if the main effect variable significantly contributed to the model.

Six econometric models were adopted: OLS, GLM, MM-estimation, censored least absolute deviation (CLAD), fractional regression model (FRM), and beta binomial (BB) regression. The OLS, which depends on a normally distributed error term, is the most commonly used regression model in mapping studies [36]. The GLM is a flexible generalisation of OLS that allows our target variable to have a non-normal error distribution [14; 36]. The log link function with Gaussian family predicting EQ-5D-5L utility fitted the data well, and hence applied in the estimation of GLM.

The MM-estimation[2] is one of the robust regression estimation methods that is used when the distribution of residual is not normal or there are some outliers that affect the model [38]. It aims to obtain estimates that have the high breakdown value, which is a common measure of the proportion of outliers that can be addressed before these observations affect the model [39], and is more efficient. Similarly, the CLAD model is a robust estimator which further take into account the censoring issue of the outcome variable [40]. It provides consistent and robust estimates under

---

[2] MM-estimation estimates the regression parameter using S-estimation, which minimise the *scale* of the residual from M-estimation and then proceed with M-estimation. The S in 'S-estimation' stands for 'scale' of the residual, M in 'M-estimation' for maximum likelihood type', and MM in' MM-estimation' stands for 'minimising M-estimation'. For detail, see Yohai [37].

rather general conditions, particularly for misspecification of errors related to heteroscedasticity and non-normality [41].

The FRM was developed to model empirical bounded dependent variables defined on [0, 1] interval that exhibit piling-up at one of the two corners [42]. Unlike other parametric methods, the important advantage of this semi-parametric FRM is that it does not make any distributional assumption about an underlying structure used to obtain the outcome variable [42; 43]. Given a vector of independent variables (X) and a dependent variable (Y), the basic assumption underlying the FRM can be summarised as:

$$E(Y \mid X) = G(X\beta), \qquad\qquad (1)$$

where G(·) is a known nonlinear function satisfying $0 \leq G(·) \leq 1$ and β is a vector of parameters to be estimated. The generalised goodness-of-functional form test proposed by Ramalho *et al.* [44] suggested that '*loglog*' is the best alternative functional form for G(.) and used as a link function. It is defined as:

$$E(Y \mid X) = \exp(-e^{X\beta}) = \exp\left[-\exp(X\beta)\right] \qquad (2)$$

Another issue in FRM is to decide whether a one- or two-part model should be used, where the discrete component (observations at one or both boundaries) is modelled as a binary or multinomial model and the rest as FRM in a two-part model [45]. In EQ-5D, since boundary values are extreme scores best described by the same mechanism as all other values in the interval, there is no strong theoretical reason to use the two-part model. Further, a P-test statistic proposed by Davidson and MacKinnon [46] and applied to this particular context by Ramalho *et al.* [43] suggested no sufficient evidence to reject the null hypothesis that the one-part model is most appropriate (p = 0.128). The FRM (with *loglog* link function) was implemented using '*frm*', a user defined Stata command [43]. The observed EQ-5D-5L utilities were firstly transformed onto a 0-1 scale before entering into the regression as the dependent variable through the following equation: $(X-X_{min})/(X_{max}-X_{min})$, in which X stands for the EQ-5D-5L utility, min/max refers to the minimum/maximum value of observed EQ-5D-5L utilities in each country-specific value set. Eventually, predictions of EQ-5D-5L utilities were back transformed to the original scale.

The BB regression is a fully parametric approach for modelling bounded data. For instance, due to its known flexibility that allows a great variety of asymmetric forms, the most popular choice to estimate equation (1) is the parametric *beta* distribution [47; 48]. The re-parameterisation of the beta regression has been detailed in Ferrari and Cribari-Neto [47]. The *logit* functional form fitted the data well and hence applied as a link function in the BB regression model for predicting the EQ-5D-5L utilities. As this parametric model is not defined at the boundary values, the outcome values should be restricted to the 0 to 1 range, excluding 0 and 1. This can be achieved by linear transformation,

[Y(N-1)+0.5]/N, following earlier literature [49; 50], where N is sample size and Y is linearly transformed EQ-5D-5L utility on 0 to 1 scale.

Goodness-of-fit measures

Following external guidance [3; 36], mean absolute error (MAE) and root mean square error (RMSE), both of which were adjusted for the degrees of freedom owing to the potential varied number of predictors in each model were firstly calculated. Both MAE and RMSE may easily be influenced by the scale of the dataset because wider scale size leads to a larger error [51]. Thus, we normalised both MAE and RMSE to the range of the measured data. Such normalised RMSE (NRMSE) and normalised MAE (NMAE) are non-dimensional and facilitate the comparison between datasets or models with different scales. Model performance was also examined by the square of the correlation coefficient between the observed and predicted values adjusted for the number of predictors in the model (adjusted-r^2). Furthermore, the degree of absolute agreement between the predicted and the observed EQ-5D-5L was assessed using Lin's concordance correlation coefficient (CCC) [52]. The CCC is popular for assessing agreement between two measures, which includes both precision (degree of scatter) and accuracy (systematic bias) components [52; 53]. A value of CCC close to unity imply a good concordance between predicted and observed measures.

In the absence of an external data, cross-validation was usually conducted using existing data to evaluate the performance of the mapping algorithms for generalisability. This study applied a *k*-fold approach, where the total sample was randomly divided into k (k=5) equally sized groups. Each time, four groups were combined as an estimation sample (i.e. 80% of the full sample) to derive mapping algorithms which were then applied to the remaining validation group (i.e. 20% of the full sample). This process was repeated five times, and hence all groups in the dataset were used for both estimation and validation sample. Finally, the normalised average of RMSE and MAE, as well as mean r^2 from the five iterations, were reported for easy comparison of the models' predictive performance.

The best fitting model with good predictive accuracy was selected primarily based on three metrics derived from both the full sample and validation sample: NRMSE, NMAE and adjusted-r^2. As there is no single gold standard criterion, these model performance indicators were equally important. Thus, the model that performs best in most of these criteria should be selected. When a disagreement is observed between them, the CCC from the full sample would be applied to determine the final preferred mapping algorithm. Eventually, this preferred model was estimated using the full sample (N=924). All statistical analyses were conducted using Stata® version 14.2 (StataCorp LP, College Station, Texas, USA) except the PCA, which was carried out in SPSS version 24 (IBM Corp., Armonk, NY, USA).

6

Results

Descriptive and conceptual overlap

Table 1 reports sample characteristics and a summary of source and target instruments. The mean (SD) value was 0.786 (0.215) for EQ-5D-5L (English value set) and 0.659 (0.222) for D-39. The average scores of the D-39 domains range from 0.535 for the 'anxiety and worry' domain to 0.781 for 'social burden'. The highest correlation (Appendix Table A1) was found between the 'energy and mobility' domain of D-39 and EQ-5D-5L utility indices: it ranges from 0.692 with Uruguayan to 0.721 with the Japanese value set. For EQ-5D-5L dimensions, the strongest correlation (0.646) was observed between the 'anxiety and worry' domain of D-39 and the 'anxiety/depression' dimension of EQ-5D-5L. For the remaining four EQ-5D-5L dimensions, higher coefficients were consistently observed with the 'energy and mobility' domain of D-39, where the highest coefficient (0.612) was observed with EQ-5D-5L 'mobility' dimension and the lowest (0.449) with 'self-care'.

Results from the PCA (Table 2) revealed six principal components when the number of components was limited to those with an eigenvalue greater than 1 and explained about 70% of the total variance. The first five principal components were the same as the D-39 domains ('diabetes control', 'energy and mobility', 'anxiety and worry', 'sexual functioning', and 'social burden'), and the sixth component was related to 'self-care'. Two EQ-5D-5L dimensions ('mobility' and 'pain/discomfort') were mainly loaded onto the same component as the D-39 items that describe activities related to 'energy and mobility'. Another two EQ-5D-5L dimensions ('self-care' and 'usual activities') were mainly loaded to the last component, in which only one D-39 item (i.e., 'having trouble caring yourself') was mainly loaded onto. The EQ-5D-5L 'anxiety/depression' dimension was loaded to the 'anxiety and worry' component, together with relevant D-39 items. None of the EQ-5D-5L dimensions were mainly loaded on the 'diabetes control', 'social burden', or 'sexual functioning' components, supporting the poor correlations identified above. In sum, the EQ-5D-5L dimensions mainly loaded on three principal components ('energy and mobility', 'anxiety and worry' and 'self-care'), which explains 17.14% of the total variance.

Model performance

Table 3a and 3b summarise the performance of models (assessed by three goodness-of-fit indicators: NRMSE, NMAE and adjusted-r^2), whilst Table 4 reports the fourth indicator (the CCC) for all countries. When the English value set was considered, the fractional regression model with *loglog* as a link function was the best performing model in terms of all four indicators: NRMSE (0.1282), NMAE (0.0914), adjusted-r^2 (52.5%), and CCC (0.691). The cross-validation approach also revealed similar

results that the FRM consistently performed best. The FRM model continued to perform well for the remaining country-specific value sets. However, both OLS and FRM equally performed well for the Dutch value set. Both models had the lowest NRMSE (0.1282) and the highest CCC (0.667) in the full sample. While OLS model provided the highest adjusted r^2 in the full sample and cross-validation, the FRM model consistently produced lower NMAE than OLS model in both full sample and cross-validation; it also produced the lowest NMAE in the full sample among all estimators. Overall, we recommend FRM for the Dutch value set.

Regarding the CCC results, the FRM revealed the highest concordance, with only the exception for the Japanese value set in which the MM-estimator produced the highest CCC (followed by CLAD and FRM), confirming the stability and consistency of our model. Except for the Japanese value set (CCC = 0.716), the MM-estimator provided generally poor concordance. The lowest concordance (CCC = 0.361) was observed for the Canadian value set.

Although mapping algorithms produce adequate mean prediction for EQ-5D-5L, this may not always be true for individual predictions. The predicted English EQ-5D-5L had a mean (SD) value of 0.786 (0.157) for the best model, which is quite similar to the observed mean. The respective 25th, 50th, and 75th percentiles of the predicted values were 0.703, 0.840, and 0.911 compared with 0.712, 0.858, and 0.937 for the observed values. These values were comparable. However, our best-fitting model over-predicted at severe health states (Figure 1). For instance, the 5th, and 10th percentiles of the predicted English EQ-5D-5L value set were 0.445, and 0.550 against 0.317, and 0.462 for the observed value set, respectively (Appendix Table A2).

Regression results

Table 5 shows the regression results for the various models when the English value set was applied. The coefficient for 'energy and mobility' (EM) was positive and statistically significant (P < 0.01) in all models. The 'anxiety and worry' (AW) domain was also a significant positive predictor of EQ-5D-5L in BB and MM models. Further, the squared-term for 'energy and mobility' (EM^2) was a significant predictor of EQ-5D-5L in OLS, GLM and CLAD models. The remaining D-39 domains were either insignificant or yield logically inconsistent signs, and hence ignored in all models. Two socio-demographic variables (age and gender) were considered. However, only age revealed statistical significance in all models except for CLAD.

Table 6 reports the best performing mapping algorithms for eight country-specific EQ-5D-5L value sets based on the model performance criteria discussed above. For each country, the optimal mapping function was consistently estimated using the FRM.

Discussion

This study developed algorithms that allow the mapping of the D-39 instrument onto eight country-specific EQ-5D-5L value sets. The EQ-5D-5L utility index was modelled as a function of the D-39 domains using six regression models. The result generally showed the FRM to be the preferred model in predicting EQ-5D-5L utility values in all country-specific value sets.

The present study also assessed the empirical performance of different statistical approaches to predict EQ-5D-5L utility values. For diabetes patients, the EQ-5D-5L were found to have a moderate degree of ceiling effect (17.4% respondents reported to be in full health according to the EQ-5D-5L descriptive system), and highly skewed to the left. Alternative regression models robust to such properties were investigated, where the FRM with *loglog link* performed best in comparison with other models under consideration for each-country-specific value sets. Yet, the regression coefficients were considerably different across countries (Table 6).

Typically, many mapping studies apply standard linear regression to predict EQ-5D utilities. However, EQ-5D has two fundamental properties. It is characterised by: i) bounded nature of utility data (between the worst and the best health states), and ii) ceiling effects (i.e., piling up of utilities at 1). Under such circumstances, the effect of predictor variables cannot be constant throughout its entire range [45]. The novelty of the FRM model is that it is non-linear and more appropriate for naturally bounded data that exhibit piling-up at one of the two corners as the case in EQ-5D [43; 45]. Furthermore, FRM is a semi-parametric approach estimated by quasi-maximum likelihood, which gives consistent estimators regardless of the true distribution of the outcome variable conditional on the predictors. This model also constrains predictions within the range determined by EQ-5D utility.

The current study differs in several important aspects from the previous study [14]. The study by Chen et al. [14] only considered two regression models (OLS and GLM). The present study, however, considered different analytical approaches, addressing the characteristics of the data such as censoring, problems of normality and heterogeneity of variance by comparing six distinct econometric models. Furthermore, while the study by Chen and colleagues used the interim 'cross-walk' vale sets, the mapping algorithm provided in this study employed the directly elicited EQ-5D-5L value sets. Thus, the preferred models and their performance in terms of both MAE and RMSE as well as adjusted-r^2 were quite different as expected. An MAE, RMSE and adjusted-r^2 values for our preferred model were 0.0914, 0.1282 and 52.5%, respectively as compared to 0.131, 0.177 and 47.5% in the study by Chen and colleagues[3]. This discrepancy may partly be attributable to variation in the target instrument used and partly due to the mapping functions employed. The two studies also differ in terms of the use of additional covariates in predicting EQ-5D-5L utility values. While the

---

[3] RMSE and MAE were adjusted for scale differences, and normalised RMSE and normalised MAE were reported in the present study.

study by Chen *et al.* [14] included country dummies and gender, the present study has considered respondents' age and gender alone. Country variable was initially considered as a potential predictor in this study. However, it does not improve the prediction of EQ-5D-5L utility and hence dropped from the analysis. In both studies, gender was not a significant predictor of EQ-5D-5L utility values.

Mapping models usually over-estimate for patients with severe health states and under-estimate for patients with good health states. This general problem of mapping studies is due to regression to the *mean* [36]. Furthermore, this might also be explained by the preference pattern of EQ-5D-5L, where substantial decrements in preference weights occur at severe health states [30]. The conceptual difference between the source and target instruments may be another explanation. From the PCA result, 'self-care' and 'usual activities' dimensions of EQ-5D-5L were mainly loaded onto the 'self-care' component, while only one D-39 item (i.e., 'having trouble caring for yourself') mainly loaded onto the same component. Many of the D-39 items related to the domains 'diabetes control', 'sexual functioning' and 'social burden', formed three distinct components unrelated to the EQ-5D-5L dimensions. These components revealed logically inconsistent signs in the estimated coefficients (results not reported here), which may be attributable to the hidden unknown interrelationships among the domains of the D-39, evidenced by relatively high correlation coefficients between the 'energy and mobility' domain with the rest of the D-39 domains.

In addition to providing insight into the dimensional structure of both D-39 and EQ-5D-5L instruments, the PCA also provided a basis for the choice between a direct or an indirect (response) mapping approach [32]. As demonstrated in the PCA (Table 2), nearly 21 items from D-39, which was mainly loaded onto the 'diabetes control', 'sexual functioning' and 'social burden' components, were not related to EQ-5D-5L dimensions. Thus, the probability of accurately predicting five EQ-5D-5L response levels for each dimension was low. In general, exact prediction of a health state requires correct predictions for each dimension of EQ-5D in response mapping approach, which is rarely achieved. Consequently, response mapping can be severely penalised when an incorrect prediction is made [54].

This study has several strengths. First, we have investigated several mapping algorithms that could be used to predict EQ-5D-5L utility when only the D-39 instrument has been applied in a clinical study. The FRM and BB models are appropriate for bounded outcome variable as the case in this study, which has the ability to predict within the given range. Despite its strong assumptions, the OLS regression model has performed pretty well as compared to other robust candidate estimators, such as GLM, BB regression, CLAD and MM-estimation. Secondly, we have adjusted for scale differences to facilitate comparison between instruments or models with different scales, which is usually undermined in many mapping studies. Thirdly, this is the first study to assess the predictive accuracy

of different EQ-5D-5L value sets using the D-39 instrument. Boland *et al.* [55] examined the clinical chronic obstructive pulmonary disease with three EQ-5D-3L value sets (for Netherland, UK and the US), though they did not make direct comparison as in the present study. Lastly, our algorithm has wider generalisability because of the multinational nature of the patient population used. Yet, as generalisability is a major issue for mapping studies, it should be tested how these models perform in different diabetic patient populations. With regard to study limitations, the data set on which the modelling is based does not include Asian subjects. If Asian diabetes patients would describe their problems differently on the source (D-39) - and the target (EQ-5D-5L) instruments, the regression coefficients in the mapping functions for China, Japan and Korea should be interpreted with some caution. Secondly, although the estimation data in this study covers a wide range of EQ-5D-5L utilities, it does not cover the whole theoretical range; consequently the prediction error towards the lower end of utility distribution tends to be large. Furthermore, self-selection bias might have occurred, as respondents were volunteered to participate in the online survey.

In conclusion, results from this study reveal that the estimated EQ-5D-5L utility values from the preferred mapping algorithm can adequately predict a mean utility value in other sample in the absence of preference-based generic instruments. Thus, such mapping algorithms facilitate health economic evaluation that can guide decision-makers for appropriate interventions in diabetic patients.

Compliance with Ethical Standards

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical Approval: Ethical approval was granted by the Monash University Human Research Ethics Committee [Reference No. CF11/ 3192–2011001748]. 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. Brazier, J., & Dixon, S. (1995). The use of condition specific outcome measures in economic appraisal. Health Economics, 4(4), 255-264. DOI: 10.1002/hec.4730040402.

2. Dolan, P. (1997). Modeling valuations for EuroQol health states. Med Care, 35. DOI: 10.1097/00005650-199711000-00002.

3. 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. DOI: https://doi.org/10.1016/j.jval.2012.10.010.

4. Brazier, J., Ratcliffe, J., Saloman, J., & Tsuchiya, A. (2017). Measuring and Valuing Health Benefits for Economic Evaluation: Oxford University Press.

5. WHO. (2016). Global Report on Diabetes. France: World Health Organization.

6. IDF. (2015). Diabetes Atlas (7th ed.). Brussels: International Diabetes Federation (IDF).

7. Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. (2015). Methods for the economic evaluation of health care programme (4th ed.). Oxford University Press: Oxford.

8. Boyer, J. G., & Earp, J. A. (1997). The development of an instrument for assessing the quality of life of people with diabetes. Diabetes-39. Med Care, 35(5), 440-453.

9. Richardson, J., McKie, J., & Bariola, E. (2014). Multi attribute utility instruments and their use. In A. J. Culyer (Ed.), Encyclopedia of health economics (pp. 341-357). San Diego: Elsevier Science.

10. Wisløff, T., Hagen, G., Hamidi, V., Movik, E., Klemp, M., & Olsen, J. A. (2014). Estimating QALY Gains in Applied Studies: A Review of Cost-Utility Analyses Published in 2010. Pharmacoeconomics, 32(4), 367-375. DOI: http://10.1007/s40273-014-0136-z.

11. Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol Group: Past, Present and Future. Applied Health Economics and Health Policy, 15(2), 127-137. DOI: 10.1007/s40258-017-0310-5.

12. NICE (National Institute for Health and Clinical Excelence). (2013). Guide to the methods of technology appraisal. London: National Health Service. Retrieved September 18, 2017, from http://www.nice.org.uk

13. Dakin, H. (2013). Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health Qual Life Outcomes., 11. DOI: 10.1186/1477-7525-11-151.

14. Chen, G., Iezzi, A., McKie, J., Khan, M. A., & Richardson, J. (2015). Diabetes and quality of life: Comparing results from utility instruments and Diabetes-39. Diabetes Research and Clinical Practice, 109(2), 326-333. DOI: http://dx.doi.org/10.1016/j.diabres.2015.05.011.

15. van Hout, B., Janssen, M. F., Feng, Y.-S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P., & Pickard, A. S. (2012). Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets. Value in Health, 15(5), 708-715. DOI: http://dx.doi.org/10.1016/j.jval.2012.02.008.

16. Augustovski, F., Rey-Ares, L., Irazola, V., Garay, O. U., Gianneo, O., Fernandez, G., Morales, M., Gibbons, L., & Ramos-Goni, J. M. (2015). An EQ-5D-5L value set based on Uruguayan population preferences. Qual Life Res. DOI: 10.1007/s11136-015-1086-4.

17. Kim, S.-H., Ahn, J., Ock, M., Shin, S., Park, J., Luo, N., & Jo, M.-W. (2016). The EQ-5D-5L valuation study in Korea. Quality of Life Research, 25(7), 1845-1852. DOI: 10.1007/s11136-015-1205-2.

18. Luo, N., Liu, G., Li, M., Guan, H., Jin, X., & Rand-Hendriksen, K. (2017). Estimating an EQ-5D-5L Value Set for China. Value in Health, 20(4), 662-669. DOI: http://doi.org/10.1016/j.jval.2016.11.016.

19. Ramos-Goni, J. M., Pinto-Prades, J. L., Oppe, M., Cabases, J. M., Serrano-Aguilar, P., & Rivero-Arias, O. (2014). Valuation and Modeling of EQ-5D-5L Health States Using a Hybrid Approach. Med Care. DOI: 10.1097/mlr.0000000000000283.

20. Versteegh, M. M., Vermeulen, K. M., Evers, S. M. A. A., de Wit, G. A., Prenger, R., & Stolk, E. A. (2016). Dutch Tariff for the Five-Level Version of EQ-5D. Value in Health. DOI: http://dx.doi.org/10.1016/j.jval.2016.01.003.

21.   Xie, F., Pullenayegum, E., Gaebel, K., Bansback, N., Bryan, S., Ohinmaa, A., Poissant, L., & Johnson, J. A. (2015). A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada. Med Care. DOI: 10.1097/mlr.0000000000000447.

22.   Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & van Hout, B. (2017). Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ. DOI: 10.1002/hec.3564.

23.   Shiroiwa, T., Ikeda, S., Noto, S., Igarashi, A., Fukuda, T., Saito, S., & Shimozuma, K. (2016). Comparison of Value Set Based on DCE and/or TTO Data: Scoring for EQ-5D-5L Health States in Japan. Value in Health, 19(5), 648-654. DOI: http://dx.doi.org/10.1016/j.jval.2016.03.1834.

24.   Petrou, S., Rivero-Arias, O., Dakin, H., Longworth, L., Oppe, M., Froud, R., & Gray, A. (2015). Preferred reporting items for studies mapping onto preference-based outcome measures: The MAPS statement. Health and Quality of Life Outcomes, 13(1), 106. DOI: 10.1186/s12955-015-0305-6.

25.   Richardson, J., Iezzi, A., & Maxwell, A. (2012). Cross-National Comparison of Twelve Quality of Life Instruments: Mic Paper 1 Background, Questions, Instruments. Research Paper 76. Retrieved November 23, 2017, from https://www.aqol.com.au/papers/researchpaper76.pdf

26.   Kaambwa, B., Chen, G., Ratcliffe, J., Iezzi, A., Maxwell, A., & Richardson, J. (2017). Mapping Between the Sydney Asthma Quality of Life Questionnaire (AQLQ-S) and Five Multi-Attribute Utility Instruments (MAUIs). Pharmacoeconomics, 35(1), 111-124. DOI: 10.1007/s40273-016-0446-4.

27.   Mihalopoulos, C., Chen, G., Iezzi, A., Khan, M. A., & Richardson, J. (2014). Assessing outcomes for cost-utility analysis in depression: comparison of five multi-attribute utility instruments with two depression-specific outcome measures. Br J Psychiatry, 205(5), 390-397. DOI: 10.1192/bjp.bp.113.136036.

28.   Chen, G., McKie, J., Khan, M. A., & Richardson, J. R. (2014). Deriving health utilities from the MacNew Heart Disease Quality of Life Questionnaire. European Journal of Cardiovascular Nursing, 14(5), 405-415. DOI: 10.1177/1474515114536096.

29.   Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res, 20(10), 1727-1736. DOI: 10.1007/s11136-011-9903-x.

30.   Olsen, J. A., Lamu, A. N., & Cairns, J. (2018). In search of a common currency: A comparison of seven EQ-5D-5L value sets. Health Econ, 27(1), 39-49. DOI: 10.1002/hec.3606.

31.   Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: how to use principal components analysis. Health Policy and Planning, 21(6), 459-468. DOI: 10.1093/heapol/czl029.

32.   Oppe, M., Devlin, N., & Black, N. (2011). Comparison of the Underlying Constructs of the EQ-5D and Oxford Hip Score: Implications for Mapping. Value in Health, 14(6), 884-891. DOI: https://doi.org/10.1016/j.jval.2011.03.003.

33.   Jobson, J. (2012). Applied Multivariate Data Analysis: Volume II: Categorical and Multivariate Methods: Springer New York.

34.   Yaremko, R. M., Harari, H., Harrison, R. C., & Lynn, E. (2013). Handbook of Research and Quantitative Methods in Psychology: For Students and Professionals: Taylor & Francis.

35.   Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4(3), 272-299. DOI: Doi 10.1037//1082-989x.4.3.272.

36.   Brazier, J. E., 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. Eur J Health Econ, 11. DOI: 10.1007/s10198-009-0168-z.

37.   Yohai, V. J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. The Annals of Statistics, 15(2), 642-656.

38.   Susanti, Y., Pratiwi, H., H., S. S., & Liana, T. (2014). M Estimation, S Estimation, and MM Estimation in Robust Regression. International Journal of Pure and Applied Mathematics, 91(3), 349-360. DOI: 10.12732/ijpam.v91i3.7.

39. Ayinde, K., Lukman, A. F., & Arowolo, O. (2015). Robust Regression Diagnostics of Influential Observations in Linear Regression Model. Open Journal of Statistics, 5(4). DOI: 10.4236/ojs.2015.54029.

40. Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. Journal of Econometrics, 25(3), 303-325. DOI: http://dx.doi.org/10.1016/0304-4076(84)90004-6.

41. Hao, L., & Naiman, D. Q. (2007). Quantile Regression: SAGE Publications.

42. Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J Appl Econometrics, 11(6), 619-632. DOI: 3.0.co;2-1.

43. Ramalho, E. A., Ramalho, J. J. S., & Murteira, J. M. R. (2011). Alternative Estimating and Testing Empirical Strategies for Fractional Regression Models. Journal of Economic Surveys, 25(1), 19-68. DOI: 10.1111/j.1467-6419.2009.00602.x.

44. Ramalho, E. A., Ramalho, J. J. S., & Murteira, J. M. R. (2014). A Generalized Goodness-of-functional Form Test for Binary and Fractional Regression Models. The Manchester School, 82(4), 488-507. DOI: 10.1111/manc.12032.

45. Ramalho, J. J. S., & da Silva, J. V. (2009). A two-part fractional regression model for the financial leverage decisions of micro, small, medium and large firms. Quantitative Finance, 9(5), 621-636. DOI: 10.1080/14697680802448777.

46. Davidson, R., & MacKinnon, J. G. (1981). Several Tests for Model Specification in the Presence of Alternative Hypotheses. Econometrica, 49(3), 781-793. DOI: 10.2307/1911522.

47. Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta Regression or Modeling Rates and Proportions. J Appl Statist, 31(7), 799-815. DOI: 10.1080/0266476042000214501.

48. Paolino, P. (2001). Maximum likelihood estimation of models with beta distributed dependent variables. Polit Anal, 9. DOI: 10.1093/oxfordjournals.pan.a004873.

49. Hunger, M., Baumert, J., & Holle, R. Analysis of SF-6D Index Data: Is Beta Regression Appropriate? Value in Health, 14(5), 759-767. DOI: 10.1016/j.jval.2010.12.009.

50. Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta distributed dependent variables. Psychol Methods, 11. DOI: 10.1037/1082-989x.11.1.54.

51. Versteegh, M. M., Leunis, A., Luime, J. J., Boggild, M., Uyl-de Groot, C. A., & Stolk, E. A. (2012). Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. Med Decis Making, 32. DOI: 10.1177/0272989x11427761.

52. Barnhart, H. X., Haber, M., & Song, J. (2002). Overall concordance correlation coefficient for evaluating agreement among multiple observers. Biometrics, 58(4), 1020-1027.

53. Andrews, G., & Slade, T. (2001). Interpreting scores on the Kessler Psychological Distress Scale (K10). Aust N Z J Public Health, 25(6), 494-497.

54. 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. DOI: 10.1177/0272989X05284108.

55. Boland, M. R. S., van Boven, J. F. M., Kocks, J. W. H., van der Molen, T., Goossens, L. M., Chavannes, N. H., & Rutten-van Mölken, M. P. M. H. (2015). Mapping the Clinical Chronic Obstructive Pulmonary Disease Questionnaire onto Generic Preference-Based EQ-5D Values. Value in Health, 18(2), 299-307. DOI: http://dx.doi.org/10.1016/j.jval.2014.11.006.

Table 1 Summary statistics (N=924)

Variables	Mean (SD)	Min	Max
EQ-5D-5L value set			
England	0.786 (0.215)	-0.157	1
Netherland	0.735 (0.255)	-0.409	1
Spain	0.764 (0.207)	-0.143	1
Canada	0.778 (0.193)	-0.044	0.949
Uruguay	0.880 (0.143)	-0.003	1
China	0.761 (0.251)	-0.252	1
Japan	0.767 (0.185)	0.101	1
Korea	0.779 (0.173)	0.112	1
Diabetes-39 total score	0.659 (0.222)	0	1
Diabetes-39 domain score			
Energy and mobility	0.628 (0.256)	0	1
Diabetes control	0.668 (0.264)	0	1
Anxiety and worry	0.535 (0.284)	0	1
Social burden	0.781 (0.242)	0	1
Sexual functioning	0.685 (0.322)	0	1
Socio-demographics			
Age	55.94 (12.579)	19	93
Gender, N (%)*			
Female	382 (41.340)	–	–
Country, N (%)*			
Australia	168 (18.180)	–	–
Canada	144 (15.580)	–	–
Germany	140 (15.150)	–	–
Norway	143 (15.480)	–	–
UK	161 (17.420)	–	–
USA	168 (18.180)	–	–

Note: The Diabetes-39 total and domain scores shown in this analysis were re-scaled onto a 0-1 scale.

* Gender and country are dummy variables.

Table 2 Principal component analysis - Rotated factor matrix

Instruments & Items	Components					
	DC	EM	AW	SF	SB	SC
Diabetes-39 items During the last month, how much was the quality of your life affected by:						
Diabetes Control (DC) domain						
1. your diabetes medication schedule	0.863					
2. following your doctor's prescribed treatment plan for diabetes	0.856					
3. food restrictions required to control your diabetes	0.874					
4. having diabetes	0.726					
5. losing control of your blood sugar levels	0.833					
6. testing your blood sugar levels	0.902					
7. the time required to control your diabetes	0.855					
8. getting your diabetes well controlled	0.845					
9. keeping a record of your blood sugar levels	0.795					
10. the need to eat at regular intervals	0.699					
11. having to organise your daily life around diabetes	0.603					
12. diabetes in general	0.579					
Energy & Mobility (EM) domain						
1. limited energy levels		0.548				
2. other health problems besides diabetes		0.705				
3. feelings of weakness		0.540				
4. restrictions on how far you can walk		0.961				
5. any daily exercises for your diabetes		0.558				
6. loss or blurring of vision		0.407				
7. not being able to do what you want		0.672				
8. other illnesses besides diabetes		0.777				
9. complications from your diabetes	0.467					
10. not being able to do housework or other jobs around the house		0.681				
11. needing to rest often		0.634				
12. problems in climbing stairs or walking up steps		0.874				
13. having trouble caring for yourself (dressing, bathing, or using the toilet)					0.427	0.508
14. restless sleep		0.443				
15. walking more slowly than others		0.816				
Anxiety and Worry (AW) domain						
1. worries about money matters			0.893			
2. concerns about your future			0.787			

	Col1	Col2	Col3	Col4	Col5	Col6
3. stress or pressure in your life			0.805			
4. feeling depressed or low			0.667			
Sexual Functioning (SF) domain						
1. diabetes interfering with your sex life				0.932		
2. problems with sexual functioning				0.990		
3. a decreased interest in sex much was the quality of				0.820		
Social Burden (SB) domain						
1. the restrictions your diabetes places on your family and friends	0.646					
2. being embarrassed because you have diabetes					0.771	
3. doing things that your family and friends don't do					[0.342]	
4. being identified as a diabetic					0.834	
5. having diabetes interfere with your family life					0.644	
EQ-5D-5L items						
1. Mobility		0.735				0.462
2. Self-care						0.811
3. Usual activities		0.542				0.569
4. Pain / discomfort		0.680				
5. Anxiety / depression			0.783			

Extraction Method: Principal Component Analysis. Pattern Matrix reported using oblique promax (Kappa=3). Loadings lower than 0.40 were suppressed, with the exception of one Diabetes-39 item in the social burden domain, in which the largest loading was reported in a bracket.

SC, self-care.

Table 3a Primary goodness-of-fit indicators of country-specific EQ-5D-5L value sets (England, Netherland, Spain, Canada)

Model	England			Netherland			Spain			Canada		
	NRMSE	NMAE	adj. r²	NRMSE	NMAE	adj. r²	NRMSE	NMAE	adj. r²	NRMSE	NMAE	adj. r²
Panel A: Full sample												
OLS	0.1288	0.0923	0.5203	**0.1282**	0.0922	**0.4976**	0.1401	0.1043	0.5406	0.1422	0.0993	0.5095
GLM	0.1301	0.0934	0.5109	0.1295	0.0939	0.4868	0.1412	0.1055	0.5334	0.1434	0.1010	0.5006
BB	0.1308	0.0952	0.5092	0.1303	0.0944	0.4823	0.1413	0.1065	0.5325	0.1454	0.1073	0.5006
FRM	**0.1282**	**0.0914**	**0.5252**	**0.1282**	0.0917	0.4972	**0.1399**	0.1043	**0.5420**	**0.1418**	0.0992	**0.5117**
MM	0.1482	0.0961	0.4829	0.1509	0.0970	0.4529	0.1405	**0.1037**	0.5407	0.1788	0.0992	0.4570
CLAD	0.1319	0.0917	0.5181	0.1299	**0.0911**	0.4860	0.1408	0.1041	0.5358	0.1435	**0.0975**	0.5075
Panel B: Cross-validation												
OLS	0.1290	0.0924	0.5176	0.1287	0.0921	**0.4979**	0.1402	0.1042	0.5370	0.1424	0.0992	0.5073
GLM	0.1302	0.0940	0.5095	0.1306	0.0945	0.4753	0.1417	0.1059	0.5264	0.1435	0.1016	0.5036
BB	0.1321	0.0955	0.5045	0.1300	0.0942	0.4770	0.1417	0.1067	0.5343	0.1458	0.1069	0.5011
FRM	**0.1281**	**0.0914**	**0.5255**	**0.1286**	**0.0918**	0.4959	**0.1400**	0.1043	**0.5433**	**0.1420**	0.0990	**0.5084**
MM	0.1474	0.0959	0.4824	0.1541	0.0983	0.4473	0.1413	0.1043	0.5397	0.1811	0.1097	0.4565
CLAD	0.1311	0.0928	0.5221	0.1307	0.0923	0.4951	0.1413	**0.1038**	0.5334	0.1447	**0.0989**	0.5075

RMSE, normalised root mean squared error; *NMAE*, normalised mean absolute error; *adj. r²*, square of correlation coefficient between predicted and observed EQ-5D-5L penalised for number of predictors; *OLS*, ordinary least square; *GLM*, generalised linear model; *BB*, binomial beta regression; *FRM*, fractional regression model; *MM*, MM-estimator; *CLAD*, censored least absolute deviation; *EQ-5D-5L*, EuroQol five-dimensional five level questionnaire. Best results are in bold type.

Table 3b Primary goodness-of-fit indicators of country-specific EQ-5D-5L value sets (Uruguay, China, Japan, and Korea)

Model	Uruguay			China			Japan			Korea		
	NRMSE	NMAE	adj. r^2	NRMSE	NMAE	adj. r^2	NRMSE	NMAE	adj. r^2	NRMSE	NMAE	adj. r^2
Panel A: Full sample												
OLS	0.1020	0.0690	0.4928	0.1361	0.0984	0.5392	**0.1378**	0.1057	0.5461	0.1392	0.1070	0.4912
GLM	0.1026	0.0697	0.4872	**0.1353**	0.1004	0.5276	0.1386	0.1065	0.5367	0.1399	0.1082	0.4858
BB	0.1035	0.0737	0.4909	0.1381	0.1022	0.5317	0.1402	0.1077	0.5356	0.1417	0.1082	0.4770
FRM	**0.1012**	0.0681	**0.5007**	**0.1353**	0.0979	**0.5444**	**0.1378**	**0.1054**	**0.5462**	**0.1388**	0.1059	**0.4941**
MM	0.1242	0.0738	0.4410	0.1387	0.0980	0.5371	0.1380	**0.1054**	0.5458	0.1495	0.1090	0.4716
CLAD	0.1033	**0.0675**	0.4906	0.1373	**0.0971**	0.5359	0.1383	0.1055	0.5439	0.1421	**0.1043**	0.4723
Panel B: Cross-validation												
OLS	0.1020	0.0691	0.4940	0.1374	0.0981	0.5318	0.1381	0.1060	**0.5450**	0.1392	0.1072	0.4879
GLM	0.1027	0.0699	0.4791	0.1375	0.1002	0.5250	0.1390	0.1066	0.5399	0.1404	0.1081	0.4859
BB	0.1033	0.0737	0.4882	0.1386	0.1023	0.5310	0.1402	0.1081	0.5359	0.1422	0.1089	0.4745
FRM	**0.1015**	**0.0682**	**0.5014**	**0.1352**	**0.0980**	**0.5460**	**0.1379**	**0.1053**	0.5448	**0.1390**	0.1058	**0.4971**
MM	0.1266	0.0747	0.4397	0.1446	0.1017	0.5354	0.1385	0.1057	0.5415	0.1568	0.1112	0.4628
CLAD	0.1024	0.0683	0.4895	0.1365	0.0983	0.5290	0.1398	0.1074	0.5412	0.1441	**0.1049**	0.4760

NRMSE, normalised root mean squared error; *NMAE*, normalised mean absolute error; *adj. r^2*, square of correlation coefficient between predicted and observed EQ-5D-5L penalised for number of predictors; *OLS*, ordinary least square; *GLM*, generalised linear model; *BB*, binomial beta regression; *FRM*, fractional regression model; *MM*, MM-estimator; *CLAD*, censored least absolute deviation; *EQ-5D-5L,* EuroQol five-dimensional five level questionnaire. Best results are in bold type.

Table 4 Goodness-of-fit: Concordance correlation coefficient (full sample)

Models	Country-specific EQ-5D-5L value sets							
	England	Netherland	Spain	Canada	Uruguay	China	Japan	Korea
OLS	0.686	**0.667**	0.703	0.676	0.662	0.702	0.708	0.660
GLM	0.675	0.653	0.695	0.666	0.655	0.688	0.702	0.654
BB	0.661	0.643	0.693	0.640	0.621	0.672	0.697	0.651
FRM	**0.691**	**0.667**	**0.706**	**0.677**	**0.669**	**0.708**	0.709	**0.664**
MM	0.495	0.432	0.698	0.361	0.381	0.677	**0.716**	0.545
CLAD	0.642	0.643	0.701	0.668	0.649	0.705	0.711	0.627

OLS, ordinary least square; *GLM*, generalised linear model; *BB*, binomial beta regression; *FRM*, fractional regression model; *MM*, MM-estimator; *CLAD*, censored least absolute deviation; *EQ-5D-5L,* EuroQol five-dimensional five level questionnaire. Best results are in bold type.

Table 5 Regression results predicting English EQ-5D-5L values from D-39 subscales

Variables	OLS	GLM	BB	FRM	MM	CLAD
EM	1.2526**	2.0272**	2.7793**	3.1014**	0.3210**	1.2708**
	(0.1367)	(0.2023)	(0.2343)	(0.1141)	(0.0312)	(0.1706)
AW			0.5296**		0.0684**	
			(0.1752)		(0.0202)	
EM^2	-0.5648**	-1.0285**				-0.6201**
	(0.1033)	(0.1427)				(0.1214)
Age	-0.0009*	-0.0009*	-0.0145**	-0.0056*	-0.0011**	
	(0.0004)	(0.0004)	(0.0033)	(0.0023)	(0.0003)	
Constant	0.3095***	-1.0116**	0.4246	0.1960	0.6617**	0.3005**
	(0.0492)	(0.0741)	(0.2403)	(0.1535)	(0.0280)	(0.0575)

Robust standard errors in parentheses. ** $p < 0.01$, * $p < 0.05$.
OLS, ordinary least square; *GLM*, generalised linear model; *BB*, binomial beta regression; *FRM*, fractional regression model; *MM*, MM-estimator; *CLAD*, censored least absolute deviation*; EM*, Energy and mobility; *AW*, anxiety and worry; EM^2, square of EM; *D-39,* Diabetes-39; *EQ-5D-5L,* EuroQol five-dimensional five level questionnaire.

Table 6 FRM as optimal mapping equations predicting each country-specific EQ-5D-5L utility values from D-39 subscales

Variables	England	Netherland	Spain	Canada	Uruguay	China	Japan	Korea
EM	3.1014**	2.7115**	2.8942**	2.6825**	3.3479**	3.3288**	2.5805**	2.3981**
	(0.1141)	(0.1464)	(0.1043)	(0.0999)	(0.1262)	(0.1156)	(0.0884)	(0.0912)
AW		0.2399*						
		(0.1415)						
Age	-0.0056*	-0.0048*	-0.0057**	-0.0051**	-0.0077**	-0.0092**	-0.0064**	-0.0062**
	(0.0023)	(0.0023)	(0.0021)	(0.0020)	(0.0026)	(0.0024)	(0.0019)	(0.0019)
Constant	0.1960	0.2154	0.0108	0.1551	0.7017**	0.2593	0.2123	0.3694**
	(0.1535)	(0.1531)	(0.1458)	(0.1347)	(0.1641)	(0.1615)	(0.1287)	(0.1248)

** $p<0.01$, * $p<0.05$. Robust standard errors in parentheses.

EM, Energy and mobility; *AW*, anxiety and worry; *D-39*, Diabetes-39; *FRM,* fractional regression model (with *loglog* link function); *EQ-5D-5L,* EuroQol five-dimensional five level questionnaire.
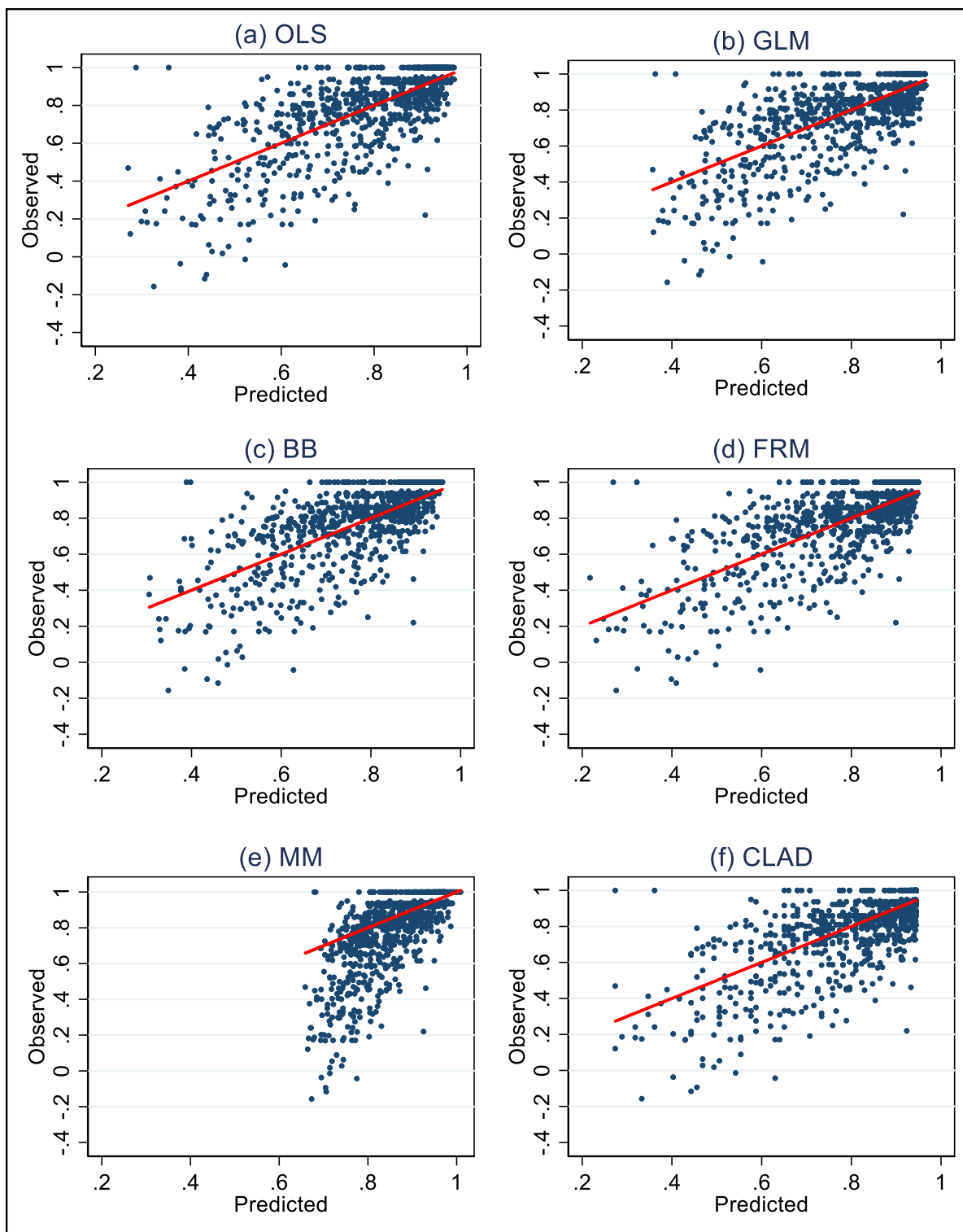
Figure 1 Scatter plot of Observed vs Predicted EQ-5D-5L (English value set).

Note: Line of equality between Observed and Predicted values (red-line); OLS, ordinary least square; *GLM*, generalised linear model; *BB*, beta binomial regression; *FRM*, fractional regression model; *MM*, robust MM-estimator; *CLAD*, censored least absolute deviation; *EQ-5D-5L*, EuroQol five-dimensional five level questionnaire.

Appendix Table A1 Spearman's rank correlations* between the EQ-5D-5L and D-39

	EM	DC	AW	SB	SF	Diabetes-39 total score
EQ-5D-5L value set						
England	0.7118	0.3604	0.4883	0.3605	0.3161	0.5516
Netherland	0.7042	0.3642	0.5153	0.3621	0.3161	0.5580
Spain	0.7172	0.3649	0.4943	0.3649	0.3194	0.5566
Canada	0.6976	0.3351	0.4463	0.3293	0.2938	0.5181
Uruguay	0.6922	0.3278	0.4341	0.3255	0.2970	0.5114
China	0.7030	0.3337	0.4474	0.3307	0.2970	0.5196
Japan	0.7209	0.3623	0.4874	0.3648	0.3233	0.5557
Korea	0.7022	0.3379	0.4434	0.3373	0.3035	0.5233
EQ-5D-5L dimensions						
Mobility	-0.6123	-0.2349	-0.2434	-0.2503	-0.2169	-0.3797
Self-Care	-0.4488	-0.2472	-0.2134	-0.2874	-0.2193	-0.3362
Usual Activities	-0.6081	-0.2848	-0.3074	-0.3016	-0.2568	-0.4290
Pain/Discomfort	-0.5670	-0.2312	-0.2970	-0.2082	-0.2109	-0.3740
Anxiety/Depression	-0.4961	-0.3729	-0.6463	-0.3484	-0.2808	-0.5277

EM, Energy and mobility; *DC*, diabetes control; *AW*, anxiety and worry; *SB,* social burden; *SF*, sexual functioning; *D-39,* Diabetes-39; *EQ-5D-5L,* EuroQol five-dimensional five level questionnaire.

* All correlation coefficients are statistically significant at $p < 0.001$ (Bonferroni-adjusted p-value).

Appendix Table A2 Mean and percentiles of observed vs predicted English EQ-5D-5L value sets

Summary statistics	Observed EQ-5D-5L	Predicted EQ-5D-5L utility					
		OLS	GLM	BB	FRM	MM	CLAD
Mean	0.786	0.786	0.787	0.780	0.786	0.845	0.800
SD	0.215	0.155	0.152	0.145	0.157	0.088	0.151
p1	0.063	0.335	0.392	0.376	0.296	0.647	0.336
p5	0.317	0.465	0.483	0.490	0.445	0.690	0.480
p10	0.462	0.549	0.548	0.561	0.550	0.723	0.573
p25	0.712	0.693	0.684	0.695	0.703	0.780	0.719
p50	0.858	0.835	0.838	0.823	0.840	0.850	0.851
p75	0.937	0.915	0.919	0.896	0.911	0.919	0.926
p90	1.000	0.937	0.936	0.926	0.934	0.958	0.941
p95	1.000	0.946	0.943	0.935	0.940	0.970	0.942
p99	1.000	0.963	0.959	0.948	0.946	0.990	0.943
iqr	0.225	0.213	0.227	0.201	0.208	0.125	0.199

p1, 1st percentile; *p5*; 5th percentile; …; *p99*, 99th percentile; *SD*, standard deviation; *iqr*, interquartile range; EQ-*5D-5L,* EuroQol five-dimensional five level questionnaire.