
Article:
Kharroubi, S.A., Beyh, Y.S., Brazier, J. orcid.org/0000-0001-8645-4780 et al. (1 more author) (2020) Modelling a preference-based index for EQ-5D-3L and EQ-5D-3L + Sleep using a Bayesian framework. Quality of Life Research, 29 (6). pp. 1495-1507. ISSN 0962-9343

https://doi.org/10.1007/s11136-020-02436-2

Modeling a preference-based index for EQ-5D-3L and EQ-5D-3L+Sleep using a Bayesian framework

Samer A Kharroubi*[1,2] , Yara S. Beyh[1], John Brazier [2] and Yaling Yang[3]

1. Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut, P.O.BOX: 11-0236, Riad El Solh 1107-2020, Beirut, Lebanon.
2. School of Health and Related Research, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA
3. Nuttfield Department of Primary Care Health Sciences, Oxford University, Oxford, United Kingdom

Abstract

Background: Conventionally, frequentist approach has been used to model health state valuation data. Recently researchers started to explore the use of Bayesian methods in this area.

Objectives: This paper presents an alternative approach to modeling health state valuation data of the EQ-5D-3L and EQ-5D-3L+Sleep descriptive systems, using a Bayesian framework and demonstrates its superiority to conventional frequentist methods.

Methods: The valuation study composed of 18 EQ-5D-3L health states and 18 EQ-5D-3L+Sleep health states valued by 160 members of the general public in South Yorkshire, UK using the time trade-off technique. Three different models were developed for EQ-5D-3L and EQ-5D-3L+Sleep accordingly using Bayesian Markov chain Monte Carlo simulation methods. Bayesian methods were applied to models fitted included a linear regression, random effect and random effect with covariates. The models are compared based on their predictive performance using mean predictions, root mean squared error (RMSE) and deviance information criterion (DIC). All analyses were performed using Bayesian Markov chain Monte Carlo simulation methods.

Results: The random effects with covariates model performs best under all criterions for the two preference-based measures, with RMSE (0.037) and DIC (637.5) for EQ-5D-3L and RMSE (0.019), DIC (416.4) for EQ-5D+Sleep. Compared with models previously estimated using frequentist approach, the Bayesian models reported in this paper provided better predictions of observed values.

Conclusion: Bayesian methods provide a better way to model EQ-5D-3L valuation data with and without a sleep 'bolt-on' and provide a more flexible in characterizing the full range of uncertainty inherent in these estimates.

Keywords: EQ-5D; preference-based health state measures; QALYs; sleep; Bayesian methods; MCMC

*Corresponding author: Samer A Kharroubi, Department of Nutrition and Food Sciences, Faculty of Agricultural and Food Sciences, American University of Beirut. P.O.BOX: 11-0236, Riad El Solh 1107-2020, Beirut, Lebanon. Tel: 961 1 350 000 Ext 4541 Email: sk157@aub.edu.lb ORCID ID: 0000-0002-2355-2719.

Background

There has been a growing use of quality-adjusted life-years (QALYs) as a measure of health in economic evaluations. It helps inform decisions about the allocation of resources between technologies and interventions competing for resources. The QALY is a measure which multiplies the quality adjustment for health by the time period endured in a given health state [1], from which the modification to the quality weight is generated using a preference-based measure. The latter consists of a classified structure to define health along with a value set denoting a utility value for every defined health state by that system.

Several preference-based measures of health-related quality of life (HRQoL) are currently available, including: the EQ-5D-3L [2], EQ-5D-5L [3], HUI2 [4] and HUI3 [5], AQoL [6], QWB [7], and the SF-6D [8], in addition to condition specific health surveys [9, 10]. The EQ-5D-3L has become the most widely used measure of health status and it is the preferred measure of HRQoL for health technology assessment in many countries, particularly in Europe [11-13], although it has also gained widespread use in North America [14], Asia and Australia [15, 16]. The EQ-5D-3L also has been used in population health studies and incorporated into many multinational clinical trials. A key problem for these measures has been the large number of unique health states that they define and the consequent need to model health state values from a valuation of a subset of possible states.

Health state values present a significant challenge for conventional statistical modelling procedures due to their nature, namely: skewed, truncated, non-continuous and hierarchical [8]. Previous statistical models of these data have met with some success in the SF-6D [8], EQ-5D-3L [17] and HUI2 [18]. In the main, these conventional statistical models used in previous analyses have been frequentist. A number of researchers have investigated the use of Bayesian methods for modelling health state preference data. For example, Kharroubi et al. [19-21] proposed a nonparametric Bayesian model and applied it to SF-6D standard gamble (SG) health state valuation data. This model has also been applied to the UK EQ-5D-3L with time trade-off (TTO) valuation data [22]. Further, it has been extended further to handle the joint US-UK EQ-5D-3L data set [23] and joint UK-Hong Kong and UK-Japan SF-6D data [24, 25], including a richer structure for covariate effects. Chan et al. [26] fitted a Bayesian model with random effects for respondents and health states to the US EQ-5D-3L valuation study, thereby estimating the uncertainty in the EQ-5D-3L scoring algorithm. Pullenayegum et al. [27, 28] also used a Bayesian approach for quantifying parameter uncertainty in EQ-5D-3L value sets and its impact on studies that use the EQ-5D-3L to measure health utility.

This paper focuses on the application of three different hierarchical Bayesian model specifications of increasing complexity to estimate health state utility values for the EQ-5D-3L with and without a sleep 'bolt-on' i.e. the EQ-5D-3L and EQ-5D-3L+Sleep. The results are compared with the conventional frequentist random effect regression models for each data set which have been reported previously [29]. This paper contributes to the understanding of the application of Bayesian approach on modelling health state valuation data.

The next section of this paper provides a brief description of the EQ-5D-3L descriptive system and the addition of the "sleep" dimension, as well as the UK EQ-5D-3L and EQ-5D-3L+Sleep valuation study and data used in this paper. It also provides the models used for the analysis of the data, including evaluation of model complexity, fit and prediction. The results of the different models are then presented. Finally, the implications of the results, including some directions for further research studies conclude the paper.

Methods

The EQ-5D-3L descriptive system

The EQ-5D-3L, a multi-dimensional questionnaire, comprises five health dimensions: mobility (MOB), self-care (SC), usual activities (UA), pain or discomfort (PAIN), and anxiety or depression (ANX). Each dimension has three levels as follows: "no problems" as level 1, "some/moderate problems" as level 2, and "extreme problems/unable to/confined to bed" as level 3; hence their combination generates 243 different health states, ranging from 11111 for full health and 33333 for the worst state [30].

For EQ-5D-3L to be used as a preference-based measure of HRQOL to calculate QALYs, every health state should be assigned with a utility value. In theory, the full health state has a utility of 1, while being dead has 0 as utility; furthermore, negative scores are given to health states judged worse that being dead. The utility values were estimated by the Measurement and Valuation of Health (MVH) group at the University of York in the UK, using a variant of the visual analog scale (VAS) and the time trade-off (TTO) techniques [31].

The EQ-5D-3L+Sleep descriptive system: a sleep "bolt-on" for the EQ-5D-3L

The EQ-5D-3L+Sleep descriptive system has a bolt-on to the generic conventional EQ-5D-3L developed by Yang and colleagues in 2014 [29] at the University of Sheffield in the UK. In this questionnaire, an additional sixth

dimension "sleep" (SL) has been incorporated using the same format of the three levels (no problems, some problems, and extreme problems with sleep) as used for the dimensions of the original EQ-5D-3L. This new six-dimensional system generates a total of 729 unique health states. The best health is 111111 and the worst health state is 333333.

Valuation study

A valuation study on selected EQ-5D-3L and EQ-5D+Sleep health states was conducted among the general population in South Yorkshire, UK, using the MVH TTO protocol. Details of the valuation study was reported elsewhere [29], and we briefly summarize it as below:

Selection of health states: The 18 EQ-5D-3L states and EQ-5D-3L+Sleep health states were identified using an orthogonal design (by applying the Orthoplan procedure of SPSS) which generated 18 states each for estimating an additive model. The best state defined by each instrument was included in both sets and was used as the upper anchor in the TTO valuation task. For each instrument, a further 17 (intermediate) states were selected for valuation, stratified into severity groups, and then randomly allocated to blocks of either 8 or 9 states (2 EQ-5D-3L blocks and 2 EQ-5D-3L+Sleep blocks). Finally, the worst health state defined by each instrument was added to each block. More detail on this is available in [29].

Selection of respondents: The sample size for the valuation survey was limited to 160 interviews due to the budget constraints [29]. The sample was selected using a 2-stage cluster random selection design. Respondents were randomly divided into 4 groups of 40, and every group was allocated 1 of the 4 health state blocks mentioned above; amounting to a total of 40 valuations per intermediate health state and 80 valuations for the pits state.

Interviews: The interviews comprised self-reporting health status using either EQ-5D-3L or EQ-5D-3L+Sleep (depending on which arm they had been allocated to), a ranking of health states defined by these measures, and a valuation exercise using the MVH TTO variant [32]. Furthermore, each questionnaire was followed by socio-demographic and economic questions. Additionally, respondents valuing EQ-5D-3L states were asked to complete the sleep dimension as part of the background questions at the end.

Study sample

The total number of valuations elicited from the 160 respondents is 1512 TTO values, with 770 and 742 values for the EQ-5D-3L and EQ-5D-3L+Sleep respectively. The TTO valuations were transformed in order to ensure that all health state values are bound on the [-1, 1] scale. At the individual level, transformed TTO values extended from -0.98 to 1.00. The number of observations, mean transformed TTO values and standard deviations, and maximum and minimum values are reported for the 2 instruments elsewhere [29]. As for the mean TTO values, those of the EQ-5D-3L extended from -0.23 (state 33333) to 0.61 (state 12312), whereas those generated from the EQ-5D-3L+Sleep ranged from -0.23 (state 33333) to 0.76 (state 211223). Concerning their standard deviations (SD), EQ-5D-3L states had an SD from 0.35 to 0.63 with a mean of 0.52, whilst the SD of the EQ-5D-3L+Sleep states ranged from 0.30 to 0.50 with a mean of 0.43. The skewness in the data is evident from the histogram shown in Figure 1A and 1B and descriptive statistics for the 770 and 742 individual transformed health state valuations for the 2 instruments. They show that negative values did occur and a large proportion of observations lie between 0.9 and 1.0, with 17% for the EQ-5D-3L and 15% for the EQ-5D-3L+Sleep. Additionally, the proportion of +1 responses is 5.8% (45/770) and 7.4% (55/742) for the EQ-5D-3L and EQ-5D-3L+Sleep respectively, whereas no valuations were observed at -1 for both measures. Further details on this analysis are provided in [29].

Modelling

Three different Bayesian models were developed. The models fitted included one linear regression (LR) model and two random effect (RE) models. The dependent variable in these models was the utility weight from the EQ-5D-3L or EQ-5D-3L+Sleep health state and the independent variables were dummies for each level above 1 for the five dimensions of the EQ-5D-3L or the six dimensions of the EQ-5D-3L+Sleep.

Model development.

The LR model predicts respondent j's valuation of health state i as:

$$y_{ij} = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_k X_{kij} + \varepsilon_{ij} \ , \qquad (1)$$

where, for $i = 1,2,\ldots n_j$ and $j = 1,2,\ldots m$, X_{ij} is the ith health state valued by respondent j and y_{ij} is the TTO score given by respondent j for that health state. Further, the β_k parameters are estimated within the linear regression model and ε_i is a zero-mean error term for the ith health state evaluated by the jth respondent.

For any given health state, note that $X_{\delta\lambda}$ will be defined as:

$X_{\delta\lambda}$ = 1 if, for this state, dimension δ is at level λ.

$X_{\delta\lambda}$ = 0 if, for this state, dimension δ is not at level λ

In total, there are 10 dummies for the EQ-5D-3L, with level λ = 1 on each dimension acting as the baseline for each dimension, and 12 for the EQ-5D-3L+Sleep.

The LR model is summarized in **Table 1**, with a conditional distribution that is defined as normal with a constant variance. This model assumes that the usual independent zero mean and constant variance random error term, with $\text{cov}(\varepsilon_{ij}\varepsilon_{i'j}) = 0$, $i \neq i'$. This implies that the 770 EQ-5D-3L observations from 80 respondents are coming from 770 respondents or the 742 EQ-5D-3L+Sleep observations from 80 respondents are coming from 742 respondents. For this model, the observed EQ-5D-3L or EQ-5D-3L+Sleep utility values were assumed to be sampled from a normal distribution, with the conditional mean (μ_{ij}) a function of each of the 10 levels of the EQ-5D-3L or 12 levels of the EQ-5D-3L+Sleep (**Table 1**).

The RE model recognises that ε_i from model (1) may not be independent of the respondent, so they are subdivided as:

$$u_j + e_{ij} \qquad (2)$$

u_i represents variations specific for each respondent, and this is presumed to be random across individual respondents; wheras e_{ij} represents the error term for the *i*th health state valuation of the *j*th individual, and this is presumed to be random across observations. Model (2) explicitly recognises that allocation of health states to individuals is random, with $\text{cov}(u_i, e_{ij}) = 0$.

For the third model, the mean of the conditional distribution was directly related to participant characteristics, including age, sex, level of education, household status, and self-reported health (in EQ-5D-3L dimensions and sleep), with the aim to capture the impact of the respondent covariates. See **Table 1** for more details on this.

Following Yaling and colleagues [29], decrements in utility from perfect health were modelled by using disutility (1-TTO) as the dependent variable. Thus, the coefficients of both levels 2 and 3 are expected to be positive, and level 3 should have larger coefficients.

Model estimation.

Bayesian methods were applied to test the model fit of the three developed models. Gibbs sampling Markov Chain Monte Carlo (MCMC) simulation methods using the WinBUGS software package were used to estimate the posterior probability distribution of each and every one of the three models [33, 34]. The WinBUGS code is available upon request from the corresponding author. For each model, a burn-in of 10,000 iterations was performed for the Gibbs sampler to reach convergence [35], which was determined by examining the convergence statistic for 2 MCMC sequences with different initial values. An additional 30,000 iterations were run for the sake of predictions and parameters estimation. The prior distributions for all the regression coefficients were specified as follows:

$$\beta_0,\ldots\beta_k \sim N(0,10^6), \quad \sigma^2 \sim \text{InverseGamma}(0.001,0.001)$$

A more thorough discussion of the choice of the non-informative prior distribution is available in [36].

Model reliability and validation.

To test the performance of the 3 models, we compared them based on the calculated root mean square error (RMSE) and the Bayesian Deviance Information Criterion (DIC) [34]. The latter is a combination of measures of both model fit and complexity, defined by

$$DIC = \bar{D} + P_D$$

where $\bar{D}$ represents the posterior mean deviance and P_D is the effective number of parameters representing model complexity. The best fitting model is defined by the minimum DIC [34].

Comparison of models

The three models are compared in terms of their coefficients with their associated 95% credible intervals (CI), as well as their predictive performance. Further, all presented models have their frequentist counterfactual and

so the best performing Bayesian model will be compared to its frequentist counterfactual [29]. Given the overall aim is to predict health state valuation; the best way to compare these models is via their predictive ability. This includes plots of predicted to actual values, calculations of the mean predictions and RMSE.

Results

Table 2 presents the coefficients for all models with their 95% CI for the two data sets, where bolded estimates have CIs excluding zero. For EQ-5D-3L, we notice that all coefficients had the expected positive sign across all three models, except for the second level of self-care. However, the CI of this coefficient includes zero so it is a weak inconsistency. Five out of the 10 coefficients on the main EQ-5D-3L domain have CIs excluding zero in all models, with larger disutilities associated with level 3 of each of mobility, self-care, usual activity, pain/discomfort and anxiety. The coefficients for the five EQ-5D-3L+sleep are broadly consistent. All coefficients had the expected positive sign apart from level 2 of anxiety/depression for the LR and RE models, and level 2 of sleep dimension for all three models. Eight coefficients of the main EQ-5D-3L+Sleep domains have credible intervals excluding zero in the LR and RE models, and seven coefficients in the RE+COV model, with larger disutility associated with level 2 and level 3 of mobility, self-care, usual activity and level 3 pain and anxiety. The 95% CI for both level 2 and level 3 of sleep dimension included zero in all models. In order to capture the impact of the respondent characteristics, covariates' age, sex, level of education, household status and self-reported health were added to the EQ-5D-3L and EQ-5D-3L+Sleep models (i.e. RE+COV). The results show that the 95% CI for all covariates coefficients included zero in the EQ-5D-3L except for three of the five age groups for the EQ-5D-3L+Sleep.

Table 3 presents the inferences for the mean health state utility values for the 18 valued health states in each of the descriptive systems. For each state, **Table 3** shows the observed sample mean health state utility and the predicted mean health state utility from the three different models for both tools. For the EQ-5D-3L descriptive system, the closest prediction of the best health state 11233 was obtained with the RE+COV model with 0.200 compared to the observed value 0.179. Further, the closest prediction of the pits state 33333 was obtained with the RE+COV model as well, with a value of -0.225, compared to the observed value -0.227. This applies to intermediate states too. For instance, the observed value of health state 13213 is 0.399, whereas the predicted values are 0.367, 0.371 and 0.402, obtained from LR, RE and RE+COV respectively. Hence, the RE+COV model, which contains all covariates, was able to generate the closest estimation. The same was found for the EQ-5D-3L+Sleep with the predictions improving

with the addition of extra parameters, namely the covariates. For instance, the observed value for health state 321111 is 0.385, whereas the predicted values are 0.368, 0.408 and 0.387, obtained from LR, RE and RE+COV, respectively.

A testing of the models' performance is also summarized in **Table 2**, where we can see that for the EQ-5D-3L, the RE+COV model scored the best RMSE and DIC with 0.037 and 637.5 respectively, as compared to the other two models; (LR: RMSE = 0.083, DIC = 1206; RE: RMSE = 0.086, DIC = 670) for the RE. Similarly, the RE+COV showed better performance than the LR and RE in the EQ-5D-3L+Sleep system by scoring the lowest RMSE and DIC with 0.019 and 416.4 respectively (LR: RMSE = 0.066, DIC = 886.8; RE: RMSE = 0.067, DIC = 427). Overall, RE+COV model was found to be the best performing Bayesian model.

Bayesian model versus frequentist

The predictive ability of these two methods of modeling data is compared in **Figures 2** and **3** for the two data sets. **Figure 2** presents the predicted mean health state values (line marked with squares) for the frequentist model (final column of **Table 2**), compared to actual values (line marked with diamonds), with health states ordered by predicted health state values. The line marked with triangles represents the errors obtained by the difference between the two valuations. **Figure 3** presents the same plots for the Bayesian model. These plots suggest that across the two measures the Bayesian model predicts the omitted data better than the frequentist model. Important differences can be seen between them from **Table 3**. For EQ-5D-3L, the frequentist model estimates the health state utility for the pits state to be -0.237 (final column of **Table 3**), even though the observed average for this state is -0.227, whereas the Bayesian model achieves a value of -0.225. For EQ-5D-3L+Sleep, the frequentist estimate was -0.212 compared with an observed mean of -0.233, but the Bayesian model gave -0.225. Overall, the predictive performance of the Bayesian model is better than that of the frequentist model, with RMSEs of 0.039 compared with 0.051, respectively, for the EQ-5D-3L and 0.023 compared with 0.031 for the EQ-5D-3L+Sleep. Across all valuations, the Bayesian model performs better for the RMSE with values of 0.037 for the Bayesian model and 0.049 for the frequentist model for the EQ-5D-3L and values of 0.019 compared with 0.041 for the EQ-5D-3L+Sleep.

Finally, a key advantage of the Bayesian method is that, in addition to generating the mean value and/or standard deviation, it provides the full distribution of the health state utility. **Figures A1** and **A2**, both available in online-only appendix at https://www.springer.com/journal/11136, show the probability distribution around the

predicted utility value for each of the 18 valued health states of the EQ-5D-3L and EQ-5D-3L+Sleep, respectively, using the LR model. From these distributions, the mean, median, standard deviation and corresponding 95% CIs can be calculated. Although it is common that parameter standard error estimates give some clue about the uncertainty of estimates, the posterior distributions capture the full range of uncertainty integrated in these utility estimates.

Discussion

This study presents alternatives to the approach of Yang and her colleagues [29] to modelling health state valuation data. These approaches consisted of three different hierarchical model specifications of increasing complexity to estimating EQ-5D-3L and EQ-5D-3L+Sleep health state utility values using Bayesian methods. Based on our analysis, the random effects with covariates model performs best under all criterions for the two preference-based measures, with RMSE of 0.037 and DIC of 637.5 for EQ-5D-3L and with RMSE (0.019), DIC (416.4) for EQ-5D+Sleep. The Bayesian model was found to provide better fit to the data when compared to its frequentist counterfactual [29], with RMSE of 0.049 for the EQ-5D-3L and a value of 0.041 for the EQ-5D-3L+Sleep. Hence, for both descriptive systems, our presented Bayesian model proved to be superior to the approach employed by [29]. However, this model will not be used to predict the values for the rest of the health states as the purpose of the value set is to estimate a societal mean value, as required by most HTA agencies who used QALYs to inform decision making, instead of different mean values for different respondents with various characteristics.

Another advantage of the Bayesian method is that it provides the full distribution of the utility values as a direct output from the modeling process rather than simply providing the mean value and/or standard deviation as is the case with the frequentist model. These posterior distributions are hugely important to capture the full range of uncertainty inherent in the health state utility estimates -- an increasingly important input to cost effectiveness analyses (CEA) for health technology assessment. On the other end, the frequentist model provides data on the uncertainty in the model parameters, but they usually do not provide estimates of the uncertainty in the health state predictions from the model [37].

Although the models using the Bayesian approach performed better than those estimated using the conventional Frequentist approach, both methods came to the same conclusion that the sleep dimension does not have an impact on valuations based on values obtained from the general population. It is possible that a sleep dimension is

useful or important for self-reported health status or even valuation for a specific group of people or patients who experience more sleep problems. This raises the issue of valuations from the general population and from the patient group, which is beyond the scope of this study. Furthermore, the sleep dimension might be more important in other countries, which suggests for future qualitative research or quantitative valuation studies in other countries. It is perhaps worth mentioning here that though the "sleep" dimension does not seem to be a useful bolt-on for the EQ-5D-3L value set, this might not be the case for the EQ-5D-5L.

This study has a number of limitations. The first is the sample size of the current study was relatively small for estimating a full model for valuing health states. This may limit the study's ability to detect important differences. Second, all models presented here traditionally assume that the error term ε_i in Equation 1 is normally distributed with constant variance (homoscedastic): $\varepsilon_i \sim N(0, \sigma^2)$. Previous work has shown that EQ-5D utility values do not exhibit constant variance. Furthermore, it follows from the results in Table 2 that there are a number of small, insignificant and inconsistent parameters in the models, in particular, those of the second levels in each dimension. Additional work should look to extend these models to account for nonconstant variance or heteroscedasticity. Finally, a reason for no impact of the Sleep dimension on utilities may be that when people value health states they rarely consider other aspects or consequences of sleeping difficulties, e.g. fatigue experienced the next day, or difficulties to concentrate on study or work, or lack of energy to perform usual activities. As there was no explanatory sentence after the word 'sleep' in the descriptive system (i.e. something similar to the list after usual activities), this may be a further limitation to the study.

In conclusion, this paper has examined three alternative models for predicting utilities for the EQ-5D-3L and EQ-5D-3L+Sleep descriptive systems. The analyses presented have demonstrated how utility data may be straightforwardly modelled using Bayesian methods, and model fit and complexity assessed using RMSE and DIC, which are straightforward to compute in a MCMC analysis. It showed that these Bayesian models produced better predictions than the previously published frequentist analyses. The Bayesian models are able to produce probability distributions as a direct output from the modeling process describing the uncertainty in the expected health state values – an increasingly important input to CEA for health technology assessment. We hope that this work will provide applied researchers with a practical set of tools to appropriately model outcomes in CEA.

Compliance with Ethical Standards

Disclosure of potential conflicts of interest:

Funding: This study was funded by the National Council for Scientific Research in Lebanon and the University Research Board at the American University of Beirut, Lebanon. During the preparing of this paper, YY was supported by the NIHR Oxford Biomedical Research Centre

Competing interests: SK and YB have no competing interests. JB and YY are members of the Euroqol Group who license the EQ-5D-3L.

Research involving Human Participants and/or Animals

Ethics approval: Ethical approval of this study was provided by the School of Health and Related Research (ScHARR) Ethics Committee at the University of Sheffield.

Informed consent. Respondents were provided with information sheets about the study and then informed consent was obtained before commencing the interview in their own home.

List of figures

List of tables

References

1- Brazier J, Ratcliffe J, Salomon JA,Tsuchyia A. Measuring and Valuing Health Benefits for Economic Evaluation (2nd Edition). Oxford: Oxford University Press, 2017.

2- Brooks, R., & Group, E. (1996). EuroQol: the current state of play. *Health policy,* **37**, 53-72.

3- Herdman, M, Gudex, C., Lloyd, A., Janssen, M. F., 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). *Quality of Life Research*, 20(10), 1727-1736.

4- Torrance, G. W., Feeny, D. H., Furlong, W. J., Barr, R. D., Zhang, Y., & Wang, Q. (1996). Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. *Medical care,* **34**, 702-722.

5- Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., and Boyle, M. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical care,* **40**, 113-128.

6- Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Annals of medicine,* **33**, 358-370.

7- Kaplan, R.M., Anderson, J.P., (1988). A general health policy model: update and application. *Health Services Research,* **23**, 203–235.

8- Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of health economics,* **21**, 271-292.

9- Revicki, D. A., Leidy, N. K., Brennan-Diemer, F., Sorensen, S., & Togias, A. (1998). Integrating patient preferences into health outcomes assessment: the multiattribute Asthma Symptom Utility Index. *Chest,* **114**, 998-1007.

10- Rowen, D., Brazier, J., Ara, R., & Zouraq, I. A. (2017). The role of condition-specific preference-based measures in health technology assessment. *PharmacoEconomics,* **35**, 33-41.

11- Rencz F, Gulacsi L, Drummond M et al (2016) EQ-5D in Central and Eastern Europe: 2000–2015. *Quality of Life Research* Int J Qual Life Aspects Treatment, Care Rehabil 25(11):2693–2710.

12- Rowen D, Azzabi Zouraq I, Chevrou-Severac H, van Hout B. (2017). International regulations and recommendations for utility data for health technology assessment. *Pharmacoeconomics*, 35:11–19.

13- Xie F, Gaebel K, Perampaladas K, Doble B, Pullenayegum E. (2014). Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Medical Decision Making*, 34:8–20.

14- Brooks R, Coons SJ, De Cock P, et al. (2003). EQ-5D in selected countries around the world. In: Brooks R, Rabin R, de Charro, F, eds. The Measurement and Valuation of Health Status Using EQ-SD: A European Perspective. Dordrecht: Kluwer Academic Publishers, 207-227.

15- Liu GG, Wu H., Li M., Gao C., Luo N. (2014). Chinese time trade-off values for EQ-5D health states. *Value Health*, 17: 597-604.

16- Viney R, Norman R, King MK, Cronin P, Street DJ, Knox S, Ratcliffe J. (2011). Time Trade-Off Derived EQ-5D Weights for Australia. *Value in Health,* 14(6): 928-936.

17- Dolan P. (1997). Modeling valuation for Euroqol health states. *Medical Care*, 35(11): 1095-108.

18- McCabe C, Stevens K, Roberts J, Brazier JE. (2005). Health state values for the HUI2 descriptive system: results from a UK Survey. *Health Economics*, 14:231-244.

19- Kharroubi SA, O'Hagan A, Brazier JE. (2005). Estimating Utilities from individual health state preference data: a nonparametric Bayesian approach. *Applied Statistics*, 54: 879-895.

20- Kharroubi SA, Brazier J, O'Hagan A, Roberts J. (2007). Modelling SF-6D health state preference data using a nonparametric Bayesian method. *Journal of Health Economics,* 26:597-612

21- Kharroubi, SA, Brazier J., McGhee S. (2013). Modelling SF-6D Hong Kong standard gamble health state preference data using a nonparametric Bayesian method. *Value in Health,* 16(6), 1032-1045.

22- Kharroubi, SA. and Abou Daher, C. (2018). Modelling a Preference-Based Index for EQ-5D using a nonparametric Bayesian method. *Quality of Life Research*, 27(11), 2841-2850.

23- Kharroubi SA, O'Hagan A, Brazier JE. (2010). A comparison of United States and United Kingdom EQ-5D health state valuations using a non-parametric Bayesian method. *Statistics in Medicine*, 29: 1622-34.

24- Kharroubi SA, Brazier J, McGhee S. (2014). A comparison of Hong Kong and United Kingdom SF-6D health states valuations using a nonparametric Bayesian method. *Value Health*, 17(4):397–405.

25- Kharroubi SA. (2015). A comparison of Japan and United Kingdom SF-6D health states valuations using a nonparametric Bayesian method. *Applied Health Economics and Health Policy*, 13(4):409–20.

26- Chan KKW, Xie F, Willan AR, Pullenayegum E. (2017). Underestimation of Variance of Predicted Health Utilities Derived from Multiattribute Utility Instruments: The Use of Multiple Imputation as a Potential Solution. *Medical Decision Making*, 37(3):262-272.

27- Pullenayegum EM, Chan KK, Xie F. (2016). Quantifying parameter uncertainty in EQ-5D-3L value sets and its impact on studies that use the EQ-5D-3L to measure health utility: a Bayesian approach. *Medical Decision Making*, 36(2):223–33.

28- Shams, S. and Pullenayegum EM. (2019). Reducing Uncertainty in EQ-5D Value Sets: The Role of Spatial Correlation. *Medical Decision Making*, 39(2):91-99.

29- Yang, Y., Brazier, J., & Tsuchiya, A. (2014). Effect of adding a sleep dimension to the EQ-5D descriptive system: a "bolt-on" experiment. *Medical Decision Making***, 34(1**), 42-53.

30- Group, T. E. (1990). EuroQol-a new facility for the measurement of health-related quality of life. *Health policy*, **16(3),** 199-208.

31- Patrick, D. L., Starks, H. E., Cain, K. C., Uhlmann, R. F., & Pearlman, R. A. (1994). Measuring preferences for health states worse than death. *Medical Decision Making*, **14(1),** 9-18.

32- Gudex, C. (1994). Time trade-off user manual: props and self-completion methods Working Papers 020cheop, Centre for Health Economics, University of York.

33- Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

34- Spiegelhatler D.J., Thomas A., Best N.G., and Lunn D. (2003). *WinBUGS User manual: Version 1.4*. MRC Biostatistics unit, Cambridge.

35- Gelman A, Rubin DB. (1992). Inference from iterative simulation using multiple sequences. *Statistical Sciences*, 7:457–72.

36- Natarajan, R., & Kass, R. E. (2000). Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, **95(449**), 227-237.

37- Kharroubi SA, Meads D, Edlin R, Browne C, McCabe C. (2015). Use of Bayesian Markov Chain Monte Carlo methods to estimate EQ-5D utility scores from EORTC QLQ data in Myeloma. *Medical Decision Making,* 35(3):351–60

Appendix

Figure A1. Probability distribution around the predicted utility value for the 18 valued health states of the EQ-5D-3L using linear regression model

Figure A2. Probability distribution around the predicted utility value for the 18 valued health states of the EQ-5D-3L+Sleep using linear regression model