Bayesian models with spatial correlation improve the precision of EQ-5D-5L value sets

Menglu Che[1], Feng Xie,[2] Stephanie Thomas[3], Eleanor Pullenayegum[4]

1. School of Public Health, Yale University, New Haven, USA 2. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Canada; 3. Sobey School of Business, Saint Mary's University, Halifax, Canada; 4. Hospital for Sick Children, Toronto, Canada

Abstract

Background: Health utilities derived from value sets for the EQ-5D-5L are commonly used in economic evaluations, however the precision of the value sets is of the same order of magnitude as reported minimum important differences (MIDs), which typically range from 0.05 to 0.1. We examined whether modelling spatial correlation among health states could improve the precision of the value sets.

Methods: Using data from 7 EQ-5D-5L valuation studies (Canada, China, Germany, Indonesia, Japan, Korea and the Netherlands) we compared the predictive precision of the published linear model, a recently proposed 8-parameter level-scale model, and two Bayesian models with spatial correlation. Predictive precision was quantified through the root mean squared error (RMSE) for out-of-sample predictions of state-level mean utilities on omitting individual states, as well as omitting blocks of states.

Results: In all seven countries, on omitting single health states, Bayesian models with spatial correlation improved upon the published linear model: the RMSEs for the originally published models were 0.060, 0.055, 0.060, 0.061, 0.039, 0.050 and 0.087 for Canada, China, Germany, Indonesia, Japan, Korea and the Netherlands respectively, and could be reduced to 0.044, 0.049, 0.051, 0.053, 0.037 0.037 and 0.086 respectively on using spatial correlation. On omitting blocks of health states, Bayesian models with spatial correlation led to smaller RMSEs in just one country, while the 8-parameter model led to smaller RMSEs in 5 of the 7 countries.

Discussion: Bayesian models incorporating spatial correlation and the 8-parameter models offer promising approaches to improving the precision of value sets for the EQ-5D-5L. The differential performance of the Bayesian models on omitting single states compared to omitting blocks of states suggests that designing valuation studies to capture more health states may further improve precision. We suggest that Bayesian models with spatial correlation and 8-parameter models be considered as candidate models when creating value sets, and that alternative designs be explored; this is vital given that the prediction errors in value sets need to be smaller than the MID of the instrument.

Introduction

Many economic evaluations rely on the value sets of multi-attribute utility instruments (MAUIs) in order to quantify quality-adjusted life years (QALYs) [1]. Value sets estimate the mean population utility for each possible health state captured by the MAUI, and are estimated through a valuation study. The accuracy of value sets varies; standard errors for state-wise mean utilities for the SF-6D range from 0.03 to 0.06 [2], while root mean square errors (RMSEs) for the EQ-5D-3L range from 0.03 to 0.28. Given that reported MIDs range from 0.03-0.04 (SF-6D) [3], 0.05-0.08 (EQ-5D-3L) [3-5] and 0.04 to 0.1 (EQ-5D-5L) [6-12], improvements to the accuracy of these value sets are desirable.

While increasing the number of respondents in a valuation study may seem an obvious solution, it may not be the best approach. Valuation studies typically use a sample sizes of 500-1000 [13, 14], and increasing sample sizes beyond 1000 has minimal impact on precision of the value sets [15, 16]. This is because the precision in the value set is driven by both the accuracy of regression parameters used in modelling valuation data and also by mis-specification in functional form, i.e. the fact that any non-saturated model for the mean health utility as a function of health state attributes is likely to be mis-specified [ref]. Increasing the sample size can improve precision of estimated regression parameters but has no effect on mis-specification of functional form. For example, in the United States valuation of the EQ-5D-3L, 84% of the parameter uncertainty in the value set was driven by mis-specification of the functional form [17].

Although increasing the number of respondents is unlikely to improve the accuracy of value sets, increasing the number of health states directly valued is more promising, as are alternative approaches to analysis. We consider each of these approaches below.

Valuation studies typically include direct valuation of a small subset of the health states represented by the instrument. For example, the protocol for the EQ-5D-5L has 86 of the 3125 health states being directly valued through time trade-off (TTO) tasks. Including more health states allows for more complex functional forms and hence has the potential for reduced model mis-specification. For example, valuation studies of the EQ-5D-3L typically include 42 or fewer health states, allowing the data to be modelled with main effects models plus some specific interaction terms (e.g. N3 [18] or I2, I3 [19]), but more complex models result in over-fitting. The Australian EQ-5D-3L valuation study included 197 health states and allowed for inclusion of many more interaction terms [20]. In work quantifying precision of a value set as a function of the sample size and states valued, Shams et al recommend valuing as many states as possible in order to improve predictive precision [16].

The typical approach to modelling valuation data is a linear regression, however this may not be optimal. In EQ-5D-5L valuation studies, which include direct valuations of 86 of the 3125 health states captured by the instrument, non-linear level-scale models have been shown to out-perform the traditional linear models in a number of countries [21], likely due to their parsimony. Bayesian approaches which include both a parametric functional form and a non-parametric model mis-specification term with spatial correlation have also been shown to out-perform

traditional regression approaches for both the US EQ-5D-3L valuation study [22] and the UK SF-6D valuation study [2].

As tools to improve the precision of value sets, alternative analytic models hold an important advantage over changes to the design of valuation studies: they can be used now, on existing valuation data at little to no additional cost. While promising, the performance of Bayesian models with spatial correlation has been examined only in isolated cases, and the models have never been compared head-to-head. The purpose of this study was to examine the performance of these models in a head-to-head comparison among seven EQ-5D-5L valuation studies, in order to gain a better understanding of their usefulness in improving the accuracy of estimated value sets. We consider their use in combination with designs featuring direct valuation of more health states in the discussion.

Methods

The EQ-5D-5L instrument

The EQ-5D-5L is a short questionnaire capturing 5 dimensions of quality life: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. For each dimension, respondents indicate one of 5 levels that best describes their health that day: no problems, slight problems, moderate problems, severe problems, extreme problems/unable to perform tasks. For a detailed description, please see [14]. The EQ-5D-5L descriptive system thus describes 3125 (5^5) health states, each of which is assigned a utility. These utilities are estimated in valuation studies; since utilities are country-specific [23], ideally each country has its own valuation study.

Data

We used data from the EQ-5D-5L valuation studies conducted in Canada [24], China [25], Germany [26], Indonesia [27], Japan [28], Korea [29], and the Netherlands [30]. While participants in these studies completed both composite time trade-off (cTTO) [14] and discrete choice experiments (DCEs), we consider TTO data only. In each country, the study design followed the valuation protocol developed by EuroQol Group [14], described briefly below.

The target sample size in each country was 1000. Each participant valued 10 health states. Participants were assigned at random to one of 10 blocks of 10 health states, where each block contained one mild health state (with only 1 dimension at level 2 and all others at level 1), the worst state (all dimensions at level 5), and the remaining states in each block were chosen to cover a range of severities. A total of 86 health states are represented among the 10 blocks.

Models

We compared the performance of the originally published model to that of the 8-parameter model, and Bayesian approaches with spatial correlation using the Shams and Kharroubi models. We describe each method briefly below.

For each study, we used the published functional form in a mixed effects model (with random intercepts for subjects). For example, for Canada, we use a model with 5 main effects using linear terms of each dimension (MO, SC, UA, PD, AD), 5 terms of each dimension indicating if it has level 4 or level 5 (MO45, SC45, UA45, PD45, AD45), and an interaction term (Num45sq) which is the square of the total number of level 4 or 5 states minus 1. For China, we use an 8-parameter model (described below), and for all other countries, we use a model with main effects only, where each dimension uses 4 dummy variables indicating if the states is in a level different from level 1 (MO2, MO3, …, MO5, SC2, …, SC5, …, AD2, … AD5) so in total we have 21 parameters including the intercept. The mixed effects models were fit using lmer from the R package lme4 [31].

We also fit 8-parameter level-scale models, as described in [21]. These models specify disutilities of being at levels 2, 3, 4 and 5, and assume that these are the same across dimensions up to a dimension-specific multiplicative constant. Specifically, letting μ_j be the mean utility for health state j, they take

$$\begin{aligned}1-\mu_j = {} & L_2\left(\beta_{MO}MO2_j+\beta_{SC}SC2_j+\beta_{UA}UA2_j+\beta_{PD}PD2_j+\beta_{AD}AD2_j\right)\\ & +L_3\left(\beta_{MO}MO3_j+\beta_{SC}SC3_j+\beta_{UA}UA3_j+\beta_{PD}PD3_j+\beta_{AD}AD3_j\right)\\ & +L_4\left(\beta_{MO}MO4_j+\beta_{SC}SC4_j+\beta_{UA}UA4_j+\beta_{PD}PD4_j+\beta_{AD}AD4_j\right)\\ & +\left(\beta_{MO}MO5_j+\beta_{SC}SC5_j+\beta_{UA}UA5_j+\beta_{PD}PD5_j+\beta_{AD}AD5\right)\end{aligned}$$

Where $\beta_{MO},\beta_{SC},\beta_{UA},\beta_{PD},\beta_{AD,}L_2,L_3,L_4$ are the parameters to be estimated. This model was fit using nonlinear least squares implemented by the nls function in R.

Turning now to the Bayesian models, both the Shams and Kharroubi models take

$$\mu_j = 1-\left(X_j\beta+\delta_j\right),$$

with $\delta_j \sim MVN(0,\Sigma)$, where the (j,k)th entry of Σ is

$$\begin{aligned}\Sigma_{jk} = \sigma_d^2 \exp\Big\{-\theta_w\big(&\left(\mathrm{MO}_j-\mathrm{MO}_k\right)^2+\left(\mathrm{SC}_j-\mathrm{SC}_k\right)^2+\left(\mathrm{UA}_j-\mathrm{UA}_k\right)^2+\left(\mathrm{PD}_j-\mathrm{PD}_k\right)^2\\ &+\left(\mathrm{AD}_j-\mathrm{AD}_k\right)^2\big)\Big\}.\end{aligned}$$

If TTO_{ij} is the utility for subject i valuing health state j, the Shams model takes $\mathrm{TTO}_{ij} \sim \mathrm{N}(\mu_j + u_i, \sigma_e^2)$, where the subject-level random intercepts are distributed as $N(0,\sigma_u^2)$. The Kharroubi model uses a multiplicative subject-level random effect, i.e. $\mathrm{TTO}_{ij} \sim \mathrm{N}(1-(1-\mu_j u_i), \sigma_e^2)$, and in our implementation we take $u_i \sim \mathrm{Gamma}(\sigma_u^{-2}, \sigma_u^{-2})$.

For comparability with the parametric approaches, the functional form (i.e. the choice of X) was the same as used in the parametric models (i.e., that reported by the valuation study in question).

We specified priors as follows. Each entry of the coefficients β followed uniform distribution on [0,1], with the exception of the coefficient corresponding to the Num45sq term in Canada, which was uniform on [0, 0.1]. The standard deviations $\sigma_d,\sigma_u,\sigma_e$ all followed uniform distributions on [0,1]. The spatial correlation parameter θ_w had a Uniform prior on [0.01,5/4]; our initial intention was

to follow [23] and use a Uniform on [0,5/4], but we found very small values of θ_w led to computational errors.

Models were fit using JAGS version 4.3.0 [32]. We ran 2000 iterations for all models, increasing the number of iterations until the Geweke diagnostic [33] as implemented in rjags [38] did not show evidence of non-convergence.

Model assessment

Fitted models were compared in terms of their predictive performance and logical consistency of the resulting value sets.

Predictive performance

The out-of-sample predictive performance of the models was assessed using the mean absolute error (MAE) and root mean squared error (RMSE); in all cases the quantities to be predicted were the state-specific mean utilities.

Out-of-sample predictive performance was computed on omitting health states, and on omitting blocks (which amounts to omitting both individuals and health states). Models were fit on a reduced dataset, then the predicted values for the held-out health states were compared to the observed values. Specifically, we computed the sample mean of each of the 86 health states from the full data, which is the average of valuations among all subjects who have valued the state. The predicted value for each state is obtained in each held-out sample, and their means are compared with the state means in the original full data.

Logical consistency

For each pair of health states where one state dominates the other, we checked whether the value set given by each model yielded logically consistent results, i.e. a predicted utility for the dominated health state that was larger than for the dominating health state. Logical consistency checks were done on the full dataset.

Results

On cross-validation omitting states, the Bayesian approaches performed best in all countries except the Netherlands (Table 1 and Figure 1). The Shams model was preferred over the Kharroubi model in five of the countries. Compared to the published functional forms, the Shams model led to reductions of 46%, 8%, 28%, 25%, 10%, 45% and 2% in MSE in Canada, China, Germany, Indonesia, Japan, Korea and the Netherlands respectively. The corresponding reductions for the Kharroubi model were 30%, 10%, 6% and 29% for Canada, Germany, Indonesia and Korea, respectively; in China, Japan and the Netherlands the Kharroubi model had

higher MSEs than the published functional form. In the Netherlands, the 8-parameter model performed best. In no country did the published functional form perform best.

On cross-validation omitting blocks, the 8-parameter models led to the smallest RMSEs in four countries, the published functional form in one, the Shams model in one country. In the remaining country, Korea, the Kharroubi model had RMSE smaller than the RMSEs for the other models.

The proportion of dominant pairs with logical inconsistencies was below 0.2% for both the Shams and Kharroubi models (Table 2). Logical inconsistencies were higher for the Kharroubi model than for the Shams model.

Discussion

We have shown that Bayesian approaches incorporating spatial correlation improve out-of-sample predictive accuracy when one state is held out at a time. Of the two Bayesian models considered, the Shams model out-performed the Kharroubi model in most cases. When whole blocks of 10 health states are held out, the Bayesian models no longer improve predictive accuracy, and the 8-paramter level-scale model is most often preferred. Both the Shams and Kharroubi models have fewer than 0.2% of dominant pairs showing logical inconsistencies.

Gains in predictive precision on using the Shams and Kharroubi models are in line with the 26% reduction in MSE for the US EQ-5D-3L valuation study reported by Shams [22] and the 23% reduction in MSE for the UK SF-6D valuation study reported by Kharroubi [2]. The superiority of the 8-parameter model over the published parametric functional forms was also in line with previous work [21] and raises the question of whether the functional forms used to create value sets are over-parameterized.

In predicting the mean utility for any given health state, the spatial correlation structure in the Bayesian models draws information from neighbouring health states. This explains why the performance of these models deteriorates on omitting blocks instead of states: on omitting blocks the state-level information is more sparse, leaving less information to draw on.

While our results show that the Bayesian models improve predictive accuracy on omitting single health states across a range of countries, our results are specific to the EQ-5D-5L. The performance of these methods in other instruments remains to be explored.

Even with the Bayesian models, the accuracy of the estimated value sets requires improvements. The Bayesian models with spatial correlation have MAEs of the same order of magnitude as reported MIDs for the EQ-5D-5L (which range from 0.05 to 0.1 [6-12]). Given our observation that the predictive performance of the Bayesian models decreases as the number of health states used for fitting the models decreases (i.e., on omitting blocks rather than states), we hypothesize that valuing more states, even if it meant fewer observations per state, would lead to better

predictive precision. Future work examining the optimal selection of states to be valued could play an important role in developing more accurate value sets.

We suggest that valuation studies examine the out-of-sample predictive performance of the traditional parametric functional forms in comparison to 8-parameter, Shams and Kharroubi models. Code for fitting the Shams, Kharroubi and 8-parameter models is available in the Appendix. Our results show that this could improve the accuracy of the resulting value sets and thus improve the accuracy of cost-utility analyses that use health utility instruments to measure utilities.

Tables

MAE	(a) Omitting States				(b)			
	Published	8-parameter	Shams	Kharroubi	Published	8-parameter	Shams	Kharroubi
Canada	0.050	0.050	**0.034**	0.040	0.051	0.049	**0.040**	0.043
China	0.041	0.041	**0.038**	0.043	0.042	**0.042**	0.047	0.052
Germany	0.048	0.043	**0.040**	0.045	0.050	**0.045**	0.048	0.056
Indonesia	0.049	0.050	**0.043**	0.048	0.049	0.051	**0.048**	0.050
Japan	0.031	0.040	**0.030**	0.040	**0.030**	0.031	0.034	0.037
Korea	0.041	0.039	**0.030**	0.035	0.041	0.040	0.041	**0.038**
Netherlands	0.070	**0.062**	0.069	0.075	0.069	**0.062**	0.067	0.077
RMSE								
Canada	0.060	0.063	**0.044**	0.050	0.060	0.063	**0.050**	0.053
China	0.051	0.051	**0.049**	0.055	0.053	**0.053**	0.060	0.066
Germany	0.060	0.055	**0.051**	0.057	0.064	**0.057**	0.064	0.070
Indonesia	0.061	0.062	**0.053**	0.059	**0.059**	0.062	0.064	0.061
Japan	0.039	**0.036**	0.037	0.049	0.039	**0.037**	0.043	0.047
Korea	0.050	0.049	**0.037**	0.042	0.051	0.052	0.050	**0.047**
Netherlands	0.087	**0.079**	0.086	0.095	0.087	**0.079**	0.086	0.101

Table 1: Mean Absolute Errors (MAE) and Root Mean Square Errors (RMSE) for all seven countries. Errors represent out-of-sample prediction errors on (a) states and (b) blocks

Country	Shams	Kharroubi
Canada	97	1436
Germany	152	424
Indonesia	562	770
Japan	0	113
Korea	31	145
Netherlands	1151	1520

Table 2: Of 756250 dominant pairs, the number of pairs of with logical inconsistency for the Shams and Kharroubi models

Figures

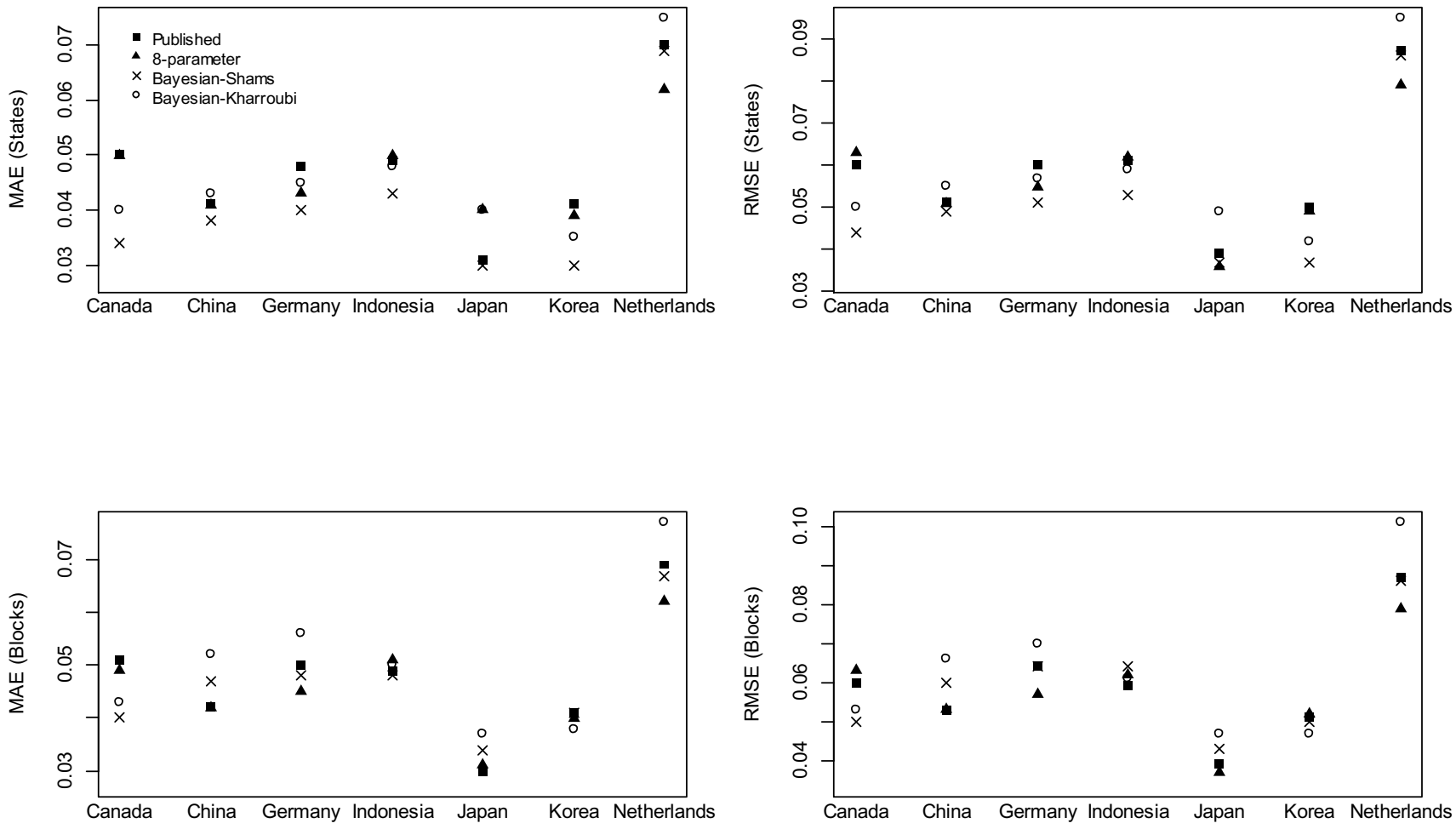

Figure 1: Root mean square errors (RMSE) and mean absolute errors (MAE) on omitting states (top row) and blocks (bottom row)

Code

Shams model

```
model{

    for(i in 1:datalength){
      mu[i] <- mu.pred[statenum86[i]] + u[idnum[i]]
      tto[i] ~ dnorm(mu[i],tausqY)
      }

# Variance-covariance matrix for the deltas
for(state1 in 1:86){
      for(state2 in 1:86){
            sigma[state1,state2] <- pow(sigmad,2)*exp(-t
      heta.w*pow(MO[state1]-MO[state2],2)
                                          - theta.w*pow(SC[state1]-
SC[state2],2)
                                          - theta.w*pow(UA[state1]-
UA[state2],2)
                                          - theta.w*pow(PD[state1]-
PD[state2],2)
                                          - theta.w*pow(AD[state1]-
AD[state2],2)
                                    )
      }
}
taud[1:86,1:86] <- inverse(sigma[,])

# compute the state-level means conditional on state level random effects
mu.pred.xbeta[1:86] <- x.mat[,]%*%beta   # this is the fixed effect component
mu.pred[1:86] ~ dmnorm(mu.pred.xbeta[],taud[,]) # this is mu (fixed + state-
level random effect)



# Subject level random effects
for(subj in 1:(num_id-1)){
      u[subj] ~ dnorm(0,tausqu)
}
u[num_id] <- -sum(u[1:(num_id-1)])   # sum to zero constraint for random
effects

# Priors

theta.w ~ dunif(0.01,1.25)
tausqu <- pow(sigmau,-2)
tausqY <- pow(sigmaY,-2)

sigmau ~ dunif(0,1)
sigmaY ~ dunif(0,1)
sigmad ~ dunif(0,1)
```

```
beta[1]~dnorm(1,1/4)
for(i in 2:21){
      beta[i] ~ dnorm(0,1/4)
}

}
```

Kharroubi model

```
model{

       for(i in 1:datalength){
      mu[i] <- 1-mu.pred[statenum86[i]]*u[idnum[i]]
      tto[i] ~ dnorm(mu[i],tausqY)
      }

# Variance-covariance matrix for the deltas
for(state1 in 1:86){
      for(state2 in 1:86){
            sigma[state1,state2] <- pow(sigmad,2)*exp(-
theta.w*pow(MO[state1]-MO[state2],2)
                                          - theta.w*pow(SC[state1]-
SC[state2],2)
                                          - theta.w*pow(UA[state1]-
UA[state2],2)
                                          - theta.w*pow(PD[state1]-
PD[state2],2)
                                          - theta.w*pow(AD[state1]-
AD[state2],2)
                                    )
      }
}
taud[1:86,1:86] <- inverse(sigma[,])

# compute the state-level means conditional on state level random effects
mu.pred.xbeta[1:86] <- x.mat[,]%*%beta   # this is the fixed effect component
mu.pred[1:86] ~ dmnorm(mu.pred.xbeta[],taud[,]) # this is mu (fixed + state-
level random effect)



# Subject level random effects
for(subj in 1:(num_id)){
   u[subj] ~ dgamma(1/tausqu,1/tausqu)
}
#u[num_id] <- -sum(u[1:(num_id-1)])   # sum to zero constraint for random
effects

# Priors
#theta.w <- 1000   # This should give correlations close to zero
theta.w ~ dunif(0.01,1.25)
tausqu <- pow(sigmau,-2)
tausqY <- pow(sigmaY,-2)
```

```
sigmau ~ dunif(0,1)
sigmaY ~ dunif(0,1)
sigmad ~ dunif(0,1)

beta[1]~dnorm(1,1/4)
for(i in 2:21){
      beta[i] ~ dnorm(0,1/4)
}
#beta[12] ~ dnorm(0,0.1)
}
```

References

1. Tosh, J.C., L.J. Longworth, and E. George, *Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals.* Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 2011. **14**(1): p. 102-109.
2. Kharroubi, S., A. O'Hagan, and J.E. Brazier, *Estimating utilities from individual health preference data: a nonparametric Bayesian method.* Applied Statistics, 2005. **54**(5): p. 879-895.
3. Walters, S. and J. Brazier, *Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D.* Quality of Life Research, 2005. **14**(6): p. 1523-1532.
4. Marra, C.A., et al., *A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis.* Soc Sci Med, 2005. **60**(7): p. 1571-82.
5. Le, Q.A., et al., *Minimal clinically important differences for the EQ-5D and QWB-SA in Post-traumatic Stress Disorder (PTSD): results from a Doubly Randomized Preference Trial (DRPT).* Health Qual Life Outcomes, 2013. **11**: p. 59.
6. Hu, X., et al., *Responsiveness and minimal clinically important difference of the EQ-5D-5L in cervical intraepithelial neoplasia: a longitudinal study.* Health Qual Life Outcomes, 2020. **18**(1): p. 324.
7. Wong, E.L.Y., R.H. Xu, and A.W.L. Cheung, *Health-related quality of life in elderly people with hypertension and the estimation of minimally important difference using EQ-5D-5L in Hong Kong SAR, China.* Eur J Health Econ, 2020. **21**(6): p. 869-879.
8. Xu, R.H., E.L. Wong, and A.W. Cheung, *Estimation of minimally important difference of the EQ-5D-5L utility scores among patients with either hypertension or diabetes or both: a cross-sectional study in Hong Kong.* BMJ Open, 2020. **10**(11): p. e039397.
9. Tsiplova, K., et al., *EQ-5D-derived health utilities and minimally important differences for chronic health conditions: 2011 Commonwealth Fund Survey of Sicker Adults in Canada.* Qual Life Res, 2016. **25**(12): p. 3009-3016.
10. Henry, E.B., et al., *Estimation of an Instrument-Defined Minimally Important Difference in EQ-5D-5L Index Scores Based on Scoring Algorithms Derived Using the EQ-VT Version 2 Valuation Protocols.* Value Health, 2020. **23**(7): p. 936-944.
11. McClure, N.S., et al., *Instrument-Defined Estimates of the Minimally Important Difference for EQ-5D-5L Index Scores.* Value Health, 2017. **20**(4): p. 644-650.
12. McClure, N.S., et al., *Small differences in EQ-5D-5L health utility scores were interpreted differently between and within respondents.* J Clin Epidemiol, 2022. **142**: p. 133-143.
13. Xie, F., et al., *Comparing EQ-5D valuation studies: A systematic review and methodological reporting checklist.* Medical Decision Making, 2013. **Mar 22 (epub ahead of print)**.
14. Oppe, M., et al., *A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol.* Value Health, 2014. **17**(4): p. 445-53.
15. Chan, K.K.W. and E.M. Pullenayegum, *The Theoretical Relationship between Sample Size and Expected Predictive Precision for EQ-5D Valuation Studies: A Mathematical Exploration and Simulation Study.* Med Decis Making, 2020. **40**(3): p. 339-347.

16. Shams, S. and E. Pullenayegum, *Design and sample size considerations for valuation studies of multi-attribute utility instruments.* Stat Med, 2020. **39**(23): p. 3074-3104.
17. Pullenayegum, E.M., K.K. Chan, and F. Xie, *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.* Med Decis Making, 2016. **36**(2): p. 223-33.
18. Dolan, P., *Modeling valuations for EuroQol health states.* Medical care, 1997. **35**(11): p. 1095-1108.
19. Shaw, J.W., J.A. Johnson, and S.J. Coons, *US valuation of the EQ-5D health states: development and testing of the D1 valuation model.* Medical care, 2005. **43**(3): p. 203-220.
20. Viney, R., et al., *Time trade-off derived EQ-5D weights for Australia.* Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 2011. **14**(6): p. 928-936.
21. Rand-Hendriksen, K., et al., *Less Is More: Cross-Validation Testing of Simplified Nonlinear Regression Model Specifications for EQ-5D-5L Health State Values.* Value in Health, 2017. **20**(7): p. 945-952.
22. Shams, S. and E. Pullenayegum, *Reducing Uncertainty in EQ-5D Value Sets: The Role of Spatial Correlation.* Med Decis Making, 2019. **39**(2): p. 91-99.
23. Norman, R., et al., *International comparisons in valuing EQ-5D health states: a review and analysis.* Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 2009. **12**(8): p. 1194-1200.
24. Xie, F., et al., *A Time Trade-off-derived Value Set of the EQ-5D-5L for Canada.* Medical Care, 2016. **54**(1): p. 98-105.
25. Luo, N., et al., *Estimating an EQ-5D-5L Value Set for China.* Value Health, 2017. **20**(4): p. 662-669.
26. Ludwig, K., J. von der Schulenburg, and W. Greiner, *German Value Set for the EQ-5D-5L.* Pharmacoeconomics, 2018. **36**(6): p. 663-674.
27. Purba, F., et al., *The Indonesian EQ-5D-5L Value Set.* Pharmacoeconomics, 2017. **35**(11): p. 1153-1165.
28. Shiroiwa, T., et al., *Comparison of Value Set Based on DCE and/or TTO Data: Scoring for EQ-5D-5L Health States in Japan.* Value Health, 2016. **19**(5): p. 648-54.
29. Kim, S.H., et al., *The EQ-5D-5L valuation study in Korea.* Qual Life Res, 2016. **25**(7): p. 1845-52.
30. Versteegh, M., et al., *Dutch Tariff for the Five-Level Version of EQ-5D.* Value in Health, 2016. **19**(4): p. 343-352.
31. Bates, D., et al., *Fitting linear mixed-effects models using lme4.* Journal of Statistical Software, 2015. **67**(1): p. 1-48.
32. Plummer, M. *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. in *International Workshop on Distributed Statistical Computing*. 2003. Vienna, Austria.
33. Geweke, J., *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, in *Bayesian Statistics*, J.M. Bernardo, et al., Editors. 1992, Oxford University Press: Oxford, U.K.
34. Plummer, M., *rjags: Bayesian Graphical Models using MCMC*. 2021.