A comparison of United States and United Kingdom EQ-5D health states valuations using a nonparametric Bayesian method

O'Hagan, A and Brazier, JE and Kharroubi, SA

The University of Sheffield

2007

HEDS Discussion Paper 07/03

ScHARR

SCHOOL OF HEALTH AND

RELATED RESEARCH

I

ScHARR

Health Economics and Decision Science Discussion Paper Series

No. 07/03

A comparison of United States and United Kingdom EQ-5D health states valuations using a nonparametric Bayesian method

O'Hagan, A.[1] Brazier, J.E.[2] Kharroubi, S.A.[3]

1. Department of Mathematics, University of York, York, UK
2. Department of Probability and Statistics, University of Sheffield, Sheffield, UK
3. School of Health and Related Research, University of Sheffield, Sheffield, UK

Corresponding author:

John E Brazier
HEDS, ScHARR, University of Sheffield
Regent Court, 30 Regent Street, Sheffield, UK, S1 4DA
Tel: +44 (0) 114 2226386
Fax: +44 (0) 114 2224095

Abstract

Few studies have compared preference values of health states obtained in different countries. This paper applies a nonparametric model to estimate and compare EQ-5D health state valuation data obtained from two countries using Bayesian methods. The data set is the US and UK EQ-5D valuation studies where a sample of 42 states defined by the EQ-5D was valued by representative samples of the general population from each country using the time trade-off technique. We estimate a function applicable across both countries which explicitly accounts for the differences between them, and is estimated using the data from both countries. The paper discusses the implications of these results for future applications of the EQ-5D and further work in this field.

1. Introduction

There has been an increasing use of preference-based measures of health related quality of life in order to calculate quality adjusted life years (QALYs) for use in cost effectiveness analyses. These preference-based measures are standardised multi-dimensional health state classifications with preference or utility weights elicited from a sample of the general population (Brazier et al, 2007a). There are currently a number of such preference-based measures, including the generic EQ-5D (Brooks, 1996), HUI2 & 3 (Torrance et al, 1996; Feeny et al, 2002), 15D (Sintonen, 1994, 1995), AQoL (Hawthorne et al, 2001), QWB (Kaplan et al, 1988) and the SF-6D (Brazier et al, 2002) and condition specific preference-based measures have also been developed (Revicki et al, 1998; Brazier et al, 2007b). These measures provide empirically derived health state value that can be used to derive QALYs for use in a cost-effectiveness analysis (Drummond et al, 2005).

The EQ-5D has become one of the most widely used measures of health status and it has been valued in the most countries of any of the generic preference-based measures. There is a concern that valuations may differ between countries due to differences in culture or health care delivery systems, thus invalidating the use of values obtained from one country in another. Earlier research with the EQ-5D found quite small and largely unimportant differences between UK, US and Spain (Johnson et al, 2005 and Badia, et al, 2001). A valuation of the HUI3 in France also found little difference with the Canadian data set. Using a much larger valuation data set obtained using the TTO from a representative sample of the US population, Johnson et al (2005) found that differences in EQ-5D health state valuations between the United States and United Kingdom general populations are potentially important. This paper extends their work by using a new nonparametric Bayesian method to model the differences between these countries that is simpler, a better fitting and more appropriate for the data that the conventional parametric of Johnson et al (2005). The objective of the study is to compare the directly measured US population-based EQ-5D health state valuations with those generated in the UK population survey and to investigate whether the valuations differ between the two populations and, if so, how.

Section 2 of this paper provides a brief description of the US and UK EQ-5D valuation surveys and the data used in this paper. Section 3 sets out a Bayesian nonparametric model that we believe provides more realistic and flexible inference

for preference functions. Section 4 presents the results of using the Bayesian method using the US/UK EQ-5D data set. We conclude with a general discussion of the results in Section 5 including some directions for future research in the merging field of cross country valuation.

2. EQ-5D data set

This instrument has a structured health state descriptive system with five dimensions of mobility, self-care, usual activities, pain/discomfort and anxiety/depression (Brooks, 1996). Each dimension has three levels of no problem (level 1), moderate or some problem (level 2) and severe problem (level 3). Together these 5 dimensions define a total of 243 health states formed by different combinations of the levels (i.e. 3^5), and each state is described in the form of a five-digit code using the three levels (e.g. state 12321 means no problems in mobility, moderate problems in self-care, and so on).

The most widely used scoring algorithm has been estimated from the valuation survey undertaken by the UK Measurement and Valuation of Health (MVH) group at York. They used a variant of the VAS and TTO in an interview survey of the UK general population (response rate 56%). Respondents were interviewed in their own home using TTO and VAS to each value 13 states. In all 43 EQ-5D states were valued in this way. Regression techniques were used to model these data to estimate additive functions with decrements for the moderate and severe dysfunctional categories of the five dimensions, a constant term for any kind of dysfunction and the term 'N3' for whenever any of the dimensions are severe

The US study used the same set of states and valued them using the same valuation methods. However, whilst the UK valuation study employed a simple sampling design, the US study used a 4-stage, cluster sampling strategy, with over sampling of the 2 largest minority groups, Hispanics and non-Hispanic blacks (Shaw et al (2005). A total of 4048 (response rate: 59.4 %) and 3395 (response rate: 64 %) respondents were interviewed in the US and UK studies, respectively. Respondents were excluded with incomplete or inconsistent responses and this resulted in usable valuation data from 3773 and 2997 respondents. These samples were found to be reasonably

representative of their populations in terms of sociodemographic characteristics (Dolan, 1997; Shaw et al, 2005).

Each respondent was asked to value 12 EQ-5D states. The US and UK studies differed in the way the 42 states were allocated across respondents. In the UK 41 health states (excluding 33333) were divided into 4 groups by severity of problems and each respondent was randomly assigned 2 "very mild," 3 "mild," 3 "moderate," 3 "severe" states, plus 33333. US respondents were randomised to get 1 of 5 groups of predefined health states: four groups were the modelling sample that each included 33333, 2 randomly selected "very mild" states (i.e. 21111, 12111, 11211, 11121 or 11112), and 9 states randomly selected from the remaining 36 states. The 5th group (i.e. the validation sample) consisted of 33333 and 11 health states randomly selected from the remaining EQ-5D states. All UK respondents were interviewed in English while in the US respondents could be interviewed in English or Spanish.

As in Johnson et al (2005), US and UK data have been pooled along with extra variables defined to reflect sampling designs of the two studies. The pooled data can is treated as coming from one study in which the population is first stratified into the US and UK sub-groups.

3. Modelling

The aim of modelling is to estimate health state utility values for all states for the EQ-5D from the 42 valued states. . The utility associated with a health state is assumed to be a function of that state, hence by estimating a relationship between the descriptive system and the observed values we can infer values for all states. Valuation surveys generate data with a complex structure creating a number of problems for estimation and a variety of techniques have been used to deal with these problems (Brazier et al, 2002). In the main these have used parametric relationships with particular assumptions about functional form, but here we contrast this parametric approach reported by Johnson et al (2005) with a more realistic and flexible nonparametric model.

A general model for health state valuations can be described by (Kharoubi et al, 2005):

$$y_j = f(\mathrm{x}_j, \alpha_j) + \varepsilon_{,j}, \qquad (1)$$

where, for $i = 1, 2, …, n_j$ and $j = 1, 2, …, m$, x_{ij} is the ith health state valued by respondent j and the dependent variable y_{ij} is the TTO valuation given by respondent j for that health state. The general model has two sets of independent random effect terms: ε_{ij} is a random error, zero-mean, term associated with each observation and α_j is a term to allow for individual characteristics of respondent j.

The interpretation of $f(\mathrm{x}_{ij}, \alpha_j)$ is as the true indifference TTO value that respondent j has for health state x_{ij}. The objective is to obtain a health state utility measure for the population as a whole, and this is generally taken to be the mean of the respondent-level health state utilities across the population. In order to account for different populations, it is possible to model α_j in terms of respondent-level covariates such as age, gender or socio-economic factors (Kharoubi et al, 2007). In the present study, we include nationality as a covariate.

3.1 The parametric approach

Johnson et al (2005) specify the following model for respondent j's health state utility:

$$f(\mathrm{x}_{ij}, \alpha_j) = \mu + \boldsymbol{\theta}' \mathrm{I}(\mathrm{x}_{ij}) + \alpha_j, \qquad (2)$$

where μ and $\boldsymbol{\theta}$ denote unknown parameters, α_j is a term to allow for individual characteristics of respondent j. Johnson et al (2005) included four terms their model (1): respondents' national identity, sex, age and age-squared.

$\mathrm{I}(\mathrm{x}_{ij})$ is a vector of dummy explanatory variables. In the simplest, no-interactions, case of this model, $\mathrm{I}(\mathrm{x}_{ij})$ is a vector of terms $I_{\delta\lambda}(\mathrm{x}_{ij})$ for each level $\lambda > 1$ of dimension δ of the EQ-5D. For example, $I_{32}(\mathrm{x}_{ij})$ denotes dimension $\delta = 3$ (usual activities), level $\lambda = 2$ (moderate problems). For any given health state x_{ij}, $I_{\delta\lambda}(\mathrm{x}_{ij})$ is defined as:

$I_{\delta\lambda}(\mathrm{x}_{ij}) = 1$ if, for state x_{ij}, dimension δ is at level λ.

$I_{\delta\lambda}(\mathrm{x}_{ij}) = 0$ if, for state x_{ij}, dimension δ is not at level λ

In all, there are 10 of these terms, with level $\lambda = 1$ acting as a baseline for each dimension. Hence the intercept parameter μ represents the health state utility value for state 111111, and summing the coefficients $\theta_{\delta\lambda}$ of the 'on' dummies derives the value of any other state.

More generally, $I(x_{,j})$ can include additional dummy variables to account for interactions between country and valuation predictors. Model (2) estimated by Johnson et al (2002) includes country-specific effects on these valuation predictors (through interaction terms) together with a constant country-specific effect across all health states (through a national identity term). This model had a further 13 variables resulting in 27 in all.

Estimation of this random effects model is via generalised least squares or maximum likelihood and so the population health state utility for state x in this model is simply $\hat{\mu} + \hat{\boldsymbol{\theta}}' I(x_{,j})$.

3.2 The nonparametric approach

The models used in all previous analyses have, like (2), been parametric. They have therefore imposed a particular form on the utility function. Kharroubi et al. (2005) propose to use a nonparametric approach that allows the function to take any form, employing Bayesian hierarchical modelling.

The Kharroubi et al. (2005) model describes the intrinsic characteristics of individual health state valuation data in a way that is argued to be more theoretically appropriate than previous parametric models. For respondent j, the health state utility of state $x_{,j}$ is

$$f(x_{,j}, \alpha_j) = 1 - \exp(\alpha_j)\left\{1 - u(x_{,j})\right\}. \qquad (3)$$

Note that the individual respondent term α_j enters multiplicatively rather than additively as in Johnson et al (2005). The term $u(x)$ is the *median* health state utility of health state x.[1] The mean health state utility is

$$\overline{u}_c(x) = 1 - \overline{\alpha}\{1 - u_c(x)\} \,,$$

where $\overline{\alpha}$ is the mean value of $\exp(\alpha)$ over the whole population. This will not in general be 1, and so the population (mean) health state utility is not the same as the median health state utility $u(x)$. More details of the nonparametric modelling and evaluation of $\overline{\alpha}$ are given in Kharroubi et al. (2005).

Kharroubi et al (2007) extended this model to allow the individual respondent effects to depend on covariates representing respondent characteristics, such as age, sex or educational status, through the further equation

$$\alpha_j \sim N(t_j^{'}\theta, \tau^2) \qquad\qquad (4)$$

where t_j is the vector of covariates for respondent *j* and θ is the vector of coefficients for the covariates. Note here that t's are centered to ensure that they have zero means, and hence that the value of $\exp(\alpha)$ for a typical person is 1. Note also that different subpopulations will have different distributions for the covariates, and hence will have different values of $\overline{\alpha}$. The model thereby provides a natural way to estimate utility functions appropriate to different subpopulations.

For the problem of interest, we wish to model specifically the particular ways in which respondents in the US value health differently from those in the UK. First, we can treat being a US or UK respondent as a covariate, using (4). This allows both populations to share an underlying utility function $u(x)$, but for US respondents to differ systematically from UK respondents in how they individually value health. Through additional interaction covariate terms, we can allow this systematic difference to depend on age, sex, or other covariates. However, it is possible that the differences between the two cultures are more subtle than this, and for instance US respondents may give lower utility values for poor health in one dimension but higher utilities when health is compromised in a different dimension. We therefore allow for

---

[1] In the Kharroubi et al (2005) model, the distribution of α_j is normal, so it has zero median as well as zero mean, and the median of $\exp(\alpha_j)$ is therefore 1.

this kind of dimension-specific difference in the two utility functions via the following amendment to the model (3):

$$f(\mathrm{x}_{.j}, \alpha_j) = 1 - \exp(\alpha_j)\left\{1 - u_c(\mathrm{x}_{.j})\right\}, \qquad (5)$$

where the term $u_c(\mathrm{x})$ in (5) is the utility of health state x in country c (c = 1 if US respondent; 0 if UK respondent), and is represented as:

$$u_0(\mathrm{x}) = \gamma_0 + \boldsymbol{\beta_0'}\,\mathrm{x} + d(\mathrm{x}), \qquad (6)$$

$$u_1(\mathrm{x}) = (\gamma_0 + \gamma_1) + (\boldsymbol{\beta_0'} + \boldsymbol{\beta_1'})\,\mathrm{x} + d(\mathrm{x}). \qquad (7)$$

The expression $\gamma_0 + \boldsymbol{\beta_0'}\,\mathrm{x}$ in (6) expresses a belief that the underlying utility function $u_0(\mathrm{x})$ for UK respondents will tend to behave like a simple linear combination of the elements of the health state description vector x. The coefficients $\boldsymbol{\beta_0}$ (which we expect to be negative) represent rates at which utility generally declines when we increase the level in the corresponding dimension of x. The comparable expression in (7) modifies these underlying trend variables with additional coefficients to reflect dimension-specific differences between the US and UK.

The term $d(\mathrm{x})$ represents a deviation from the simple linear trend that is common to both UK and US respondents. As in Kharroubi et al. (2005), $d(\mathrm{x})$ is treated as an unknown function and in a Bayesian nonparametric framework it therefore becomes a random variable. It has zero mean, constant variance and is constrained by a correlation between $d(\mathrm{x})$ and $d(\mathrm{x}')$ for two different states x and x′ which decreases as the distance between x and x′ increases. The effect of this is to assert that if x and x′ describe very similar health states (in the sense that their levels are the same or close in all dimensions) their utilities will be approximately the same, and so the preference function varies smoothly as the health state changes.

Note that the inclusion of a US/UK covariate in (4) means that the median value of $\exp(\alpha)$ in given country will not in general be one and so, unlike (3), $u_c(\mathrm{x})$ is not the *median* health state utility[2].

---

[2] Here the distribution of α_j is defined by (4), so its median is not zero, and the median of $\exp(\alpha_j)$ is therefore different than 1. This shifts the median utility in both models.

4. Results

For this analysis the vector of covariates is set to be (US, Age, Age2, Sex, US*Age, US*Age2, US*Sex) where US is a dummy variable to differentiate respondents' national identity (1 if US respondent; 0 if UK respondent) and Age2 is squared age. To further investigate how valuations differ between the two countries, the interaction terms US*Age, US*Age2 and US*Sex are included. Table 1 shows the posterior mean and standard deviation for population mean utility of the 43 health states in the US and UK. US mean valuations are higher than the UK valuations for all heath states. Predicted mean valuations for the 43 EQ-5D health states ranged from -0.2671(33333) to 1(11111) for the US population and from -0.5613 (33333) to 1(11111) for the UK population. This difference is also obvious from Figure 1 which shows the predicted mean health state valuations (line marked with diamonds) for the US population along with the predicted mean health state valuations (line marked with squares) for the UK population. The line marked with triangles represents the difference between the two valuations. As can be seen, the US and UK valuations are very similar for mild health states; whereas for intermediate and severe health states the US values are higher than the UK ones. The (US-UK) differences in the posterior means for the true population mean utilities ranging from 0 (11111) to 0.2942 (33333). Finally, posterior means for the true population mean valuations of 9 health states are less than zero for the US whereas negative posterior mean valuations are present for 17 health states for the UK population. We compare these results with the results of Johnson et al (2005) later in this section.

Figure 2 shows the posterior distributions of the covariates (US, Age, Age2, and Sex). As can be seen, these distributions are concentrated away from zero which indicates that these covariates have important effects. Note that the age (and age squared) and sex effects are consistent with those previously found for SF-6D (Kharroubi et al, 2007). The main additional finding here is that the US covariate has a substantial effect. Figure 3 shows the posterior distributions of the interaction terms US*Age, US*Age2 and US*Sex. There appears to be a substantial interaction between US/UK and sex. The magnitude is about 0.05, whereas the marginal sex effect is about -0.04 (Figure 2). Given the marginal sex effect and interaction effect are of opposite signs and of comparable size, this means that there is little net effect from the sex variable

9

in the US. Figure 4 shows the mean TTO utility values by age for the pits state (33333) for both the US and UK. We see that US posterior means for the true population mean valuations for both males and females are higher than the UK valuations for this health state. We also see that the line marked with squares is very close to the one marked with diamonds which also indicates that the sex effect in the US is small. The big difference between the line marked with triangles and the one marked with stars shows an appreciable sex effect in the UK. Finally, the age effect seems larger in the UK but this does not appear to be a substantial interaction (between US/UK and age) as the magnitude is about 0.004 (Figure 3), whereas the marginal age effect is about -0.004 (Figure 2).

Figure 5 shows the posterior distributions of the underlying regression parameters in (6) i.e. γ_0 and $\boldsymbol{\beta_0}$. The first element is the coefficient of the constant term γ_0, but the other 5 elements represent slopes of $\boldsymbol{\beta_0}$ as each of the 5 dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) increases. We see that, with high probability, all of these 5 coefficients are negative, so the fitted parametric relationship for the UK satisfies monotonicity. Figure 6 shows the posterior distributions of the elements of the corresponding γ_1 and $\boldsymbol{\beta_1}$ in (7). The first element is the change γ_1 in the constant term between the UK and US and the other 5 elements represent changes in slope $\boldsymbol{\beta_1}$ as each of the 5 dimensions increases. We see that all these parameters have posterior distributions strongly to one side of the zero value, indicating that there are genuine differences between the US and UK in how individuals respond to poor health in each of the five dimensions. Note, however, that for the first two dimensions (Mobility and Self Care) the differences are negative, indicating that US respondents are more sensitive to poor health in these dimensions, whereas positive differences in the other dimensions suggest that the US respondents were less concerned about Usual Activities, Pain/Discomfort or Anxiety/Depression than their UK counterparts. We also note, however, that the magnitudes of these parameters are much smaller than the corresponding $\boldsymbol{\beta_0}$ parameters and so this indicates rather small differences between UK and US.

Figure 7 presents the predicted mean health state valuations (line marked with diamonds) for the US population obtained by Johnson et al (2005), along with the predicted mean health state valuations (line marked with squares) for the UK population. The line marked with triangles represents the difference between the two

10

valuations. In comparison with Figure 1, Figure 7 shows more disordered behaviour. Estimated mean valuations for the 43 EQ-5D health states ranged from -0.38(33333) to 1(11111) for the US population and from -54 (33333) to 1(11111) for the UK population. The (US-UK) differences in population means range from -0.01 (21111) to 0.25 (21232). In addition, mean valuations of 8 health states are less than 0 for the US whereas negative mean valuations are present for 16 health states for the UK population. We found that the maximum difference between the two approaches is 0.1109, for pits health state, whereas the difference on average is 0.039 (Table 1).

Figure 8 presents the relationship between the US and UK predicted posterior means for the true population mean valuations of the 43 EQ-5D health states. The solid reference line is a line of unity. This figure shows that the magnitude of the differences in US and UK population mean valuations increased with decreasing valuations of these health states. In comparison with Figure 8, Figure 9 presents the corresponding relationship obtained by Johnson et al (2005).

As always, it is important to check the validity of the assumed model (3). Figure 10 plots a histogram of standardised residuals across all 81240 health state valuations for the Bayesian model. According to this model we would expect these to have approximately a standard normal distribution. Figure 10 broadly supports this, although there is some evidence of skewness. This is not surprising, given the negative skewness in the original TTO data at the individual level. However, the degree of skewness is probably not high enough to invalidate the analysis in our model, which assumes normally distributed errors.

A better test of the validity of the model is to investigate its ability to predict the values for states that have not been used in the estimation. Data relating to 3 selected health states (5204 observations) were removed randomly from the estimation data, and the models fitted on data for the remaining 40 states (76036 observations). Table 3 presents the true sample means for the 3 omitted states, together with their predicted mean and standard deviation values from the nonparametric model estimated on the reduced data set. The predictive performance of the nonparametric model is better than the parametric model overall, with root mean square error (RMSE) of 0.040 and 0.168, respectively. It can be seen that the nonparametric model predicts the omitted data quite well. It is to be noted that the predictive standard deviations here are larger

than those in Table 1, because the model in Table 1 is predicting the data on which it was estimated, whereas the model in Table 3 is predicting out of sample data.

5. Discussion

In this paper we have developed a Bayesian nonparametric model for estimating the utility values of health states defined by the EQ-5D generic descriptive system., in order to generate QALYs and hence to conduct cost utility analysis of health care interventions. We believe our model is simpler, a better fit, predicts validly and more meaningfully than the previous analysis of Johnson et al (2005).

The most important difference between our analysis and that of Johnson et al (2005) is the way that our model identifies distinct, interpretable aspects of the differences between the UK and US population valuations of EQ-5D. The dominant effect is the US covariate, whose positive value means that US respondents generally place higher utility on all the health states. Because of the multiplicative way that covariates enter our model, this fits the data better since the difference between the two countries is negligible for good health states and at its greatest for the poorest health state (33333). Johnson et al found the same effect, but only through estimating a very large number of parameters, whereas our model clearly shows this through a single dominant term. It is this parameter that can be seen in the steady trend of the line marked with triangles in Figure 1 or the steady deviation of the points from the solid line in Figure 8.

This factor is moderated through a second covariate term, the interaction between US and Sex. We see that this overall tendency to give higher utilities is stronger for females than for males, such that whereas there is a clear sex difference in the UK this effectively disappears in US respondents. Finally, we have seen that the US respondents are more sensitive to poor health in mobility and self care, but less sensitive in the dimensions of usual activities, pain and anxiety. These differences are small but emerge clearly in the analysis.

In contrast, Johnson et al (2005) model national/cultural differences through fitting separate US and UK parameters for every level (above 1) of every dimension. Through estimating many more parameters, their analysis is susceptible to over-fitting and data artefacts (which we believe can be seen in the more disordered behaviour in Figures 7 and 9 i.e. the non steady trend of the line marked with triangles in Figure 7

and the non steady deviation of the points from the solid line in Figure 9). It is also more difficult to interpret.

The novel part of the analysis was to use the covariate framework of the Bayesian model to represent the differences between the two countries. Understanding this will be hugely important in terms of reducing the need for instruments to be valued separately in each country. It is shown that, although the model gives satisfactory results for the US and UK separate fits, the combined analysis is better and so the use of the interactions terms as covariates is an acceptable model. This implies that the US and UK should use the combined data rather than their own ones.

In conclusion, we found meaningful country-specific differences in directly elicited TTO valuations of EQ-5D health states between the US and UK general populations. The differences in valuations depend on the degree of severity of the health state. The simple idea of a covariate to represent the US and UK is adequate and hugely important in terms of reducing the need for EQ-5D to be valued separately in each country. The model is applicable to other generic health state descriptive systems such as SF-6D and HUI-II, as well as to more specialised, disease-specific systems.

References

Badia, X., Roset, M., Herdman, M. and Kind, P. (2001). A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Medical Decision Making* 20, 7—16.

Brazier J.E., Ratcliffe J., Tsuchiya A., Solomon J. (2007a). Measuring and valuing health for economic evaluation. Oxford: Oxford University Press.

Brazier J.E., Czoski-Murray C., Roberts J., Brown M., Symonds T., Kelleher C.. (2007b). Estimation of a preference-based index from a condition specific measure: the King's Health Questionnaire. *Medical Decision Making* (in press)

Brazier, J.E., 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).

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

Dolan, P., 1997. Modeling valuation for Euroqol health states. *Medical Care* 35, 351(363).

Drummond, M.F., Sculpher M., O'Brien B., Stoddart, G.L., Torrance, G.W. (2005). Methods for the economic evaluation of health care programmes. Oxford: Oxford Medical Publications.

Feeny, D. H., Furlong, W. J., Torrance, G. W., Goldsmith, C.H., Zenglong, Z., Depauw, S., Denton, M., and Boyle, M., 2002. Multi-attribute and single-attribute utility function for the Health Utility Index Mark 3 system. *Medical care* 40(20), 113(128).

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

Johnson, J.A., Luo, N., Shaw, J.W., Kind, P. and Coons, S.J. (2005). Valuations of EQ-5D Health States; Are the United States and United Kingdom Different. *Medical Care* 43; 221—228.

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

Kharroubi, S. A., O'Hagan, A. and Brazier, J. E., 2005. Estimating Utilities from individual health state preference data: a nonparametric Bayesian approach. Applied Statistics 54, 879--895.

Kharroubi, S. A., O'Hagan, A. and Brazier, J. E., 2007. Modelling covariates for the SF-6D standard gamble health state preference data using a nonparametric Bayesian method. To appear in *Social Science and Medicine*.

Revicki D.A., Leidy N.K., Brennan-Diemer F., Sorenson S., Togias A. (1998). Integrating patients' preferences into health outcomes assessment: the multiattribute asthma symptom utility index. *Chest* 114(4): 998-1007.

Shaw, J.W., Johnson, J.A. and Coons, S.J., 2005. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical Care* 43, 203-220.

Sintonen H., 1994. The 15D-measure of health-related quality of life. I. Reliability, validity and sensitivity of its health state descriptive system. National Centre for Health Program Evaluation, Working Paper 41, Melbourne.

Sintonen H., 1995. The 15D-measure of health-related quality of life. II. Feasibility, reliability and validity of its valuation system. National Centre for Health Program Evaluation, Working Paper 42, Melbourne.

Torrance, G. W., Feeny, D. H., Furlong, W. J. Barr, R. D., Zhang, Y., Wang, Q. A. 1996. Multi-attribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. *Medical Care* 34(7); 702-722

Table 1: Predicted statistics for 42 health states in the US and UK General population

| Health State | US | | UK | | | Shaw's | Diff of |
	Mean	S. D.	Mean	S. D.	Difference	Difference	Diff
11111	1	0	1	0	0	0	0
11121	0.8689	0.0076	0.8146	0.0089	0.0543	0.03	0.0243
21111	0.8708	0.0071	0.859	0.0087	0.0118	-0.01	0.0218
11211	0.8702	0.0067	0.8379	0.0084	0.0323	-0.004	0.0363
12111	0.8278	0.0076	0.8104	0.0088	0.0174	0.005	0.0124
11112	0.8375	0.0066	0.783	0.0083	0.0545	-0.002	0.0565
12121	0.7648	0.0095	0.7062	0.0113	0.0586	0.04	0.0186
12211	0.769	0.0088	0.7328	0.0105	0.0362	0.01	0.0262
11122	0.7613	0.0094	0.6633	0.0117	0.098	0.03	0.068
22121	0.6698	0.0091	0.6069	0.0114	0.0629	0.08	-0.0171
22112	0.6748	0.0099	0.6175	0.0117	0.0573	0.03	0.0273
22122	0.6045	0.0107	0.5048	0.0126	0.0997	0.12	-0.0203
21222	0.6151	0.0092	0.4953	0.0127	0.1198	0.1	0.0198
12222	0.6091	0.0089	0.4897	0.0121	0.1194	0.09	0.0294
11312	0.5987	0.0097	0.4866	0.0119	0.1121	0.07	0.0421
21312	0.5699	0.0102	0.4644	0.0126	0.1055	0.07	0.0355
22222	0.5463	0.0103	0.428	0.0139	0.1183	0.07	0.0483
11113	0.483	0.0095	0.344	0.0119	0.139	0.11	0.029
13212	0.4168	0.0097	0.3094	0.0126	0.1074	0.07	0.0374
13311	0.3594	0.0097	0.2594	0.0126	0.1	0.05	0.05
12223	0.3574	0.0107	0.1706	0.0145	0.1868	0.18	0.0068
21232	0.231	0.0114	0.0171	0.0155	0.2139	0.25	-0.0361
21323	0.2714	0.0118	0.0599	0.0164	0.2115	0.14	0.0715
23321	0.2341	0.0106	0.0939	0.0153	0.1402	0.14	0.0002
11131	0.2909	0.0102	0.1105	0.0136	0.1804	0.08	0.1004
22323	0.1889	0.0116	-0.0234	0.0174	0.2123	0.22	-0.0077
32211	0.1864	0.0107	0.0769	0.0141	0.1095	0.06	0.0495
22331	0.108	0.011	-0.0966	0.0174	0.2046	0.17	0.0346
11133	0.1432	0.0107	-0.1135	0.0156	0.2567	0.2	0.0567
21133	0.1057	0.0114	-0.1459	0.0175	0.2516	0.2	0.0516
23313	0.0767	0.0111	-0.1105	0.0165	0.1872	0.14	0.0472
23232	0.0448	0.0126	-0.1743	0.0184	0.2191	0.15	0.0691
22233	0.0452	0.0131	-0.2125	0.0184	0.2577	0.19	0.0677
33212	0.0313	0.0115	-0.117	0.0161	0.1483	0.05	0.0983
32223	-0.0089	0.0127	-0.2334	0.0184	0.2245	0.19	0.0345
13332	-0.0375	0.0121	-0.2906	0.0196	0.2531	0.19	0.0631
32313	-0.0302	0.0122	-0.2364	0.0175	0.2062	0.11	0.0962
33321	-0.0473	0.0116	-0.2225	0.0176	0.1752	0.07	0.1052
32232	-0.0395	0.0137	-0.2739	0.0187	0.2344	0.16	0.0744
32331	-0.1149	0.0134	-0.3449	0.0197	0.23	0.13	0.1
33232	-0.1446	0.0143	-0.3836	0.0206	0.239	0.15	0.089
33323	-0.1845	0.0134	-0.4342	0.0214	0.2497	0.16	0.0897
33333	-0.2671	0.0122	-0.5613	0.0217	0.2942	0.17	0.1242

Figure 1. Predicted mean health state valuations for the US and UK populations using the Bayesian model
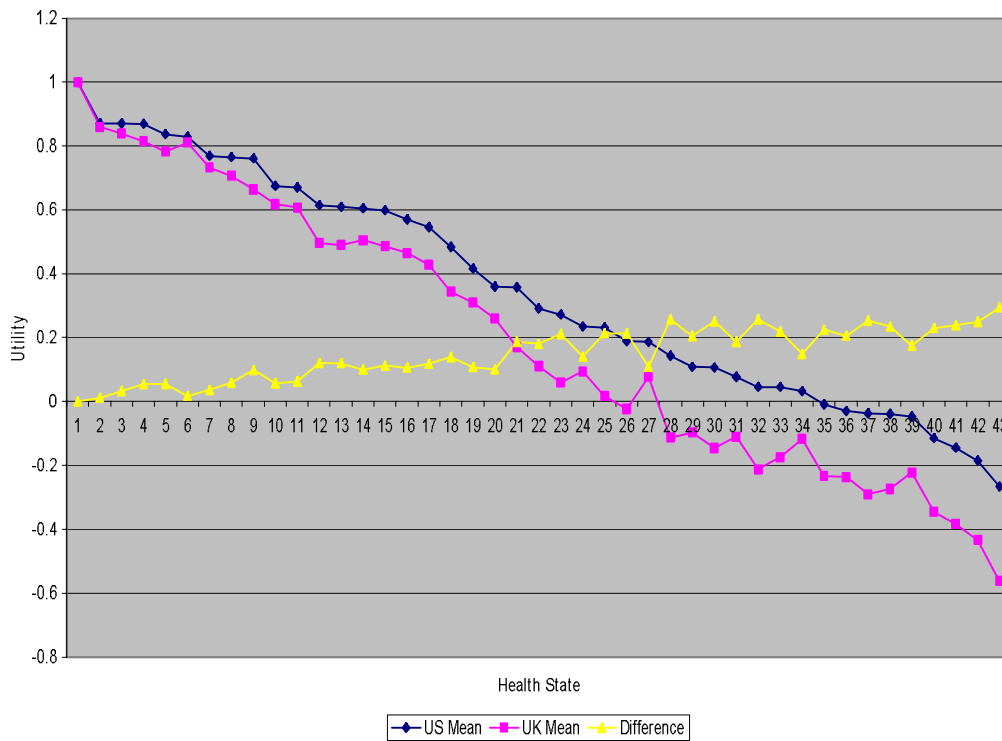

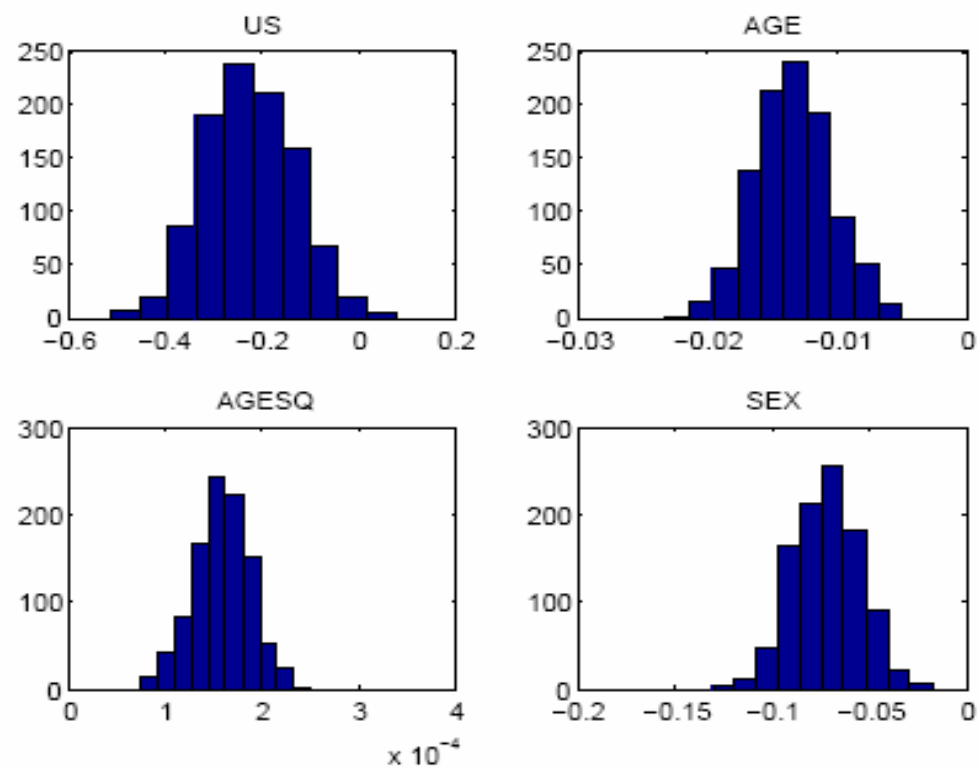Figure 2. Posterior distribution of the covariates

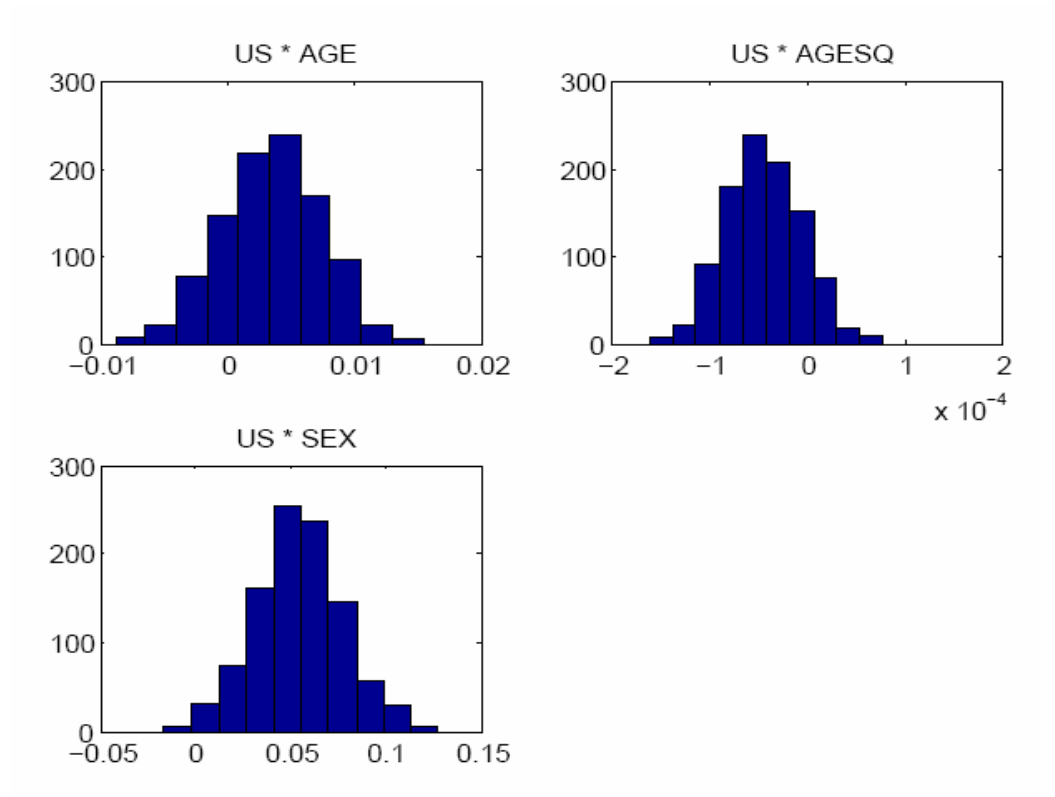Figure 3. Posterior distribution of the interaction terms

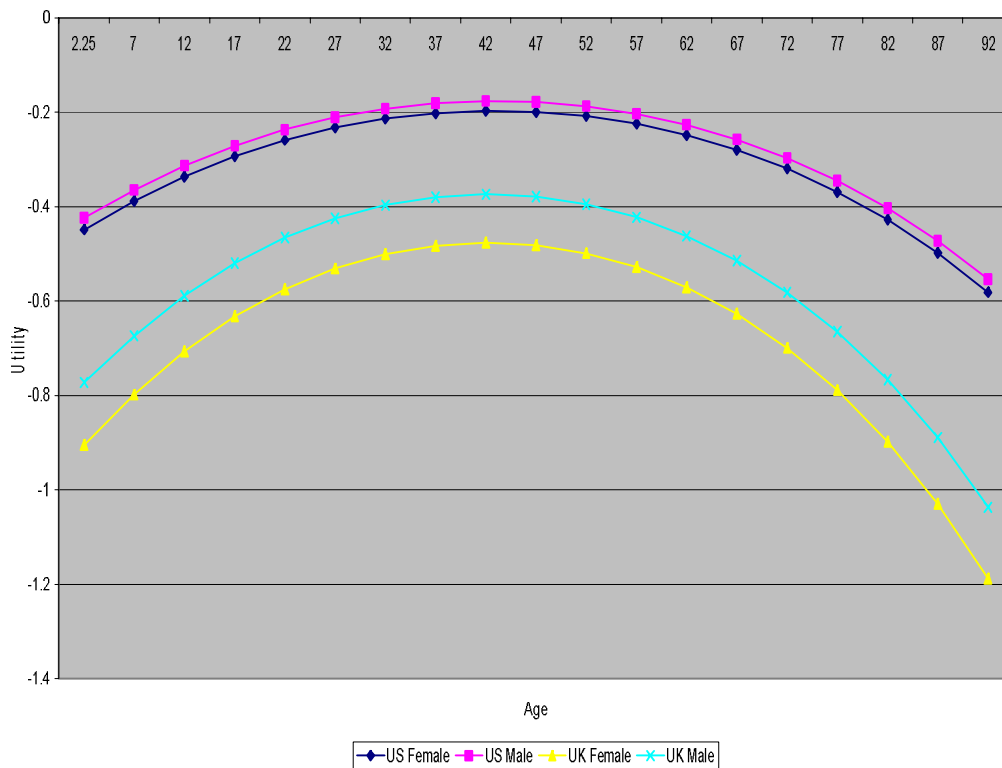Figure 4. Mean TTO utility values by age for pits state: US and UK

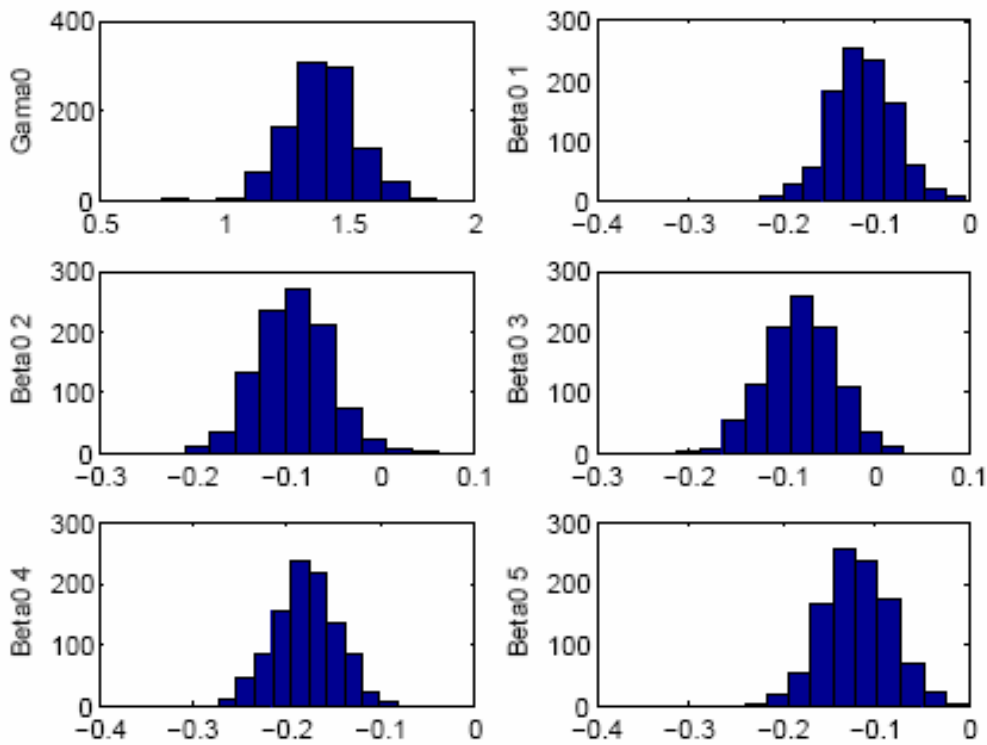Figure 5. Posterior distribution of the underlying regression parameters γ_0 and $\boldsymbol{\beta_0}$.

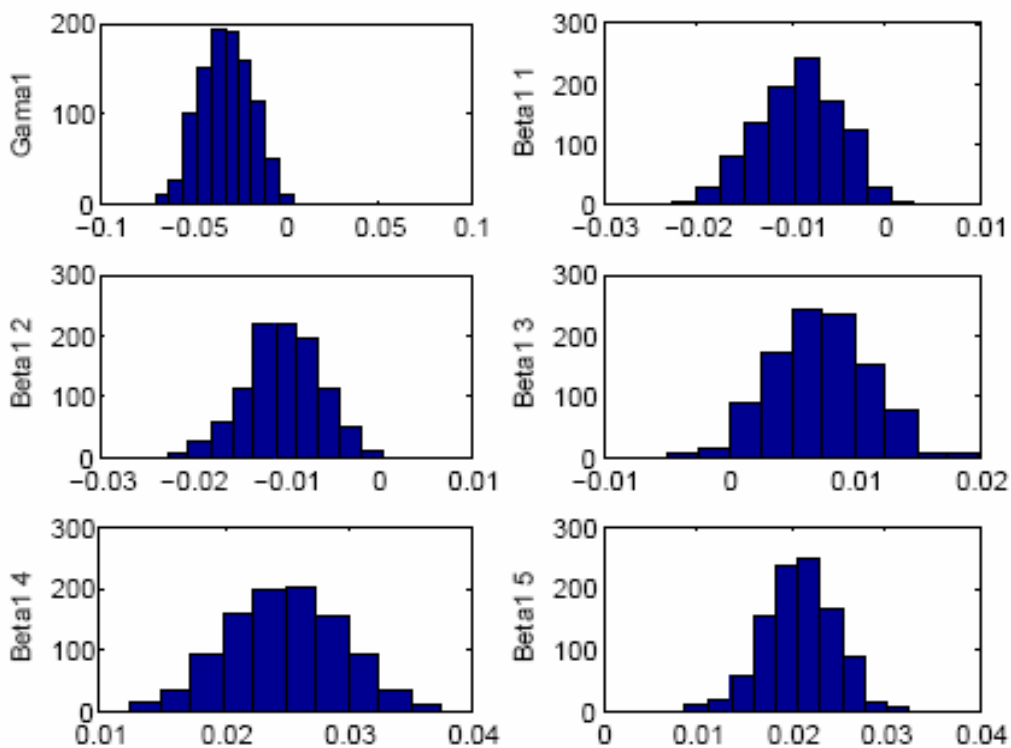Figure 6. Posterior distributions of the corresponding γ_1 and $\boldsymbol{\beta_1}$

Figure 7. Mean valuations for the US and UK populations (Johnson et al model)
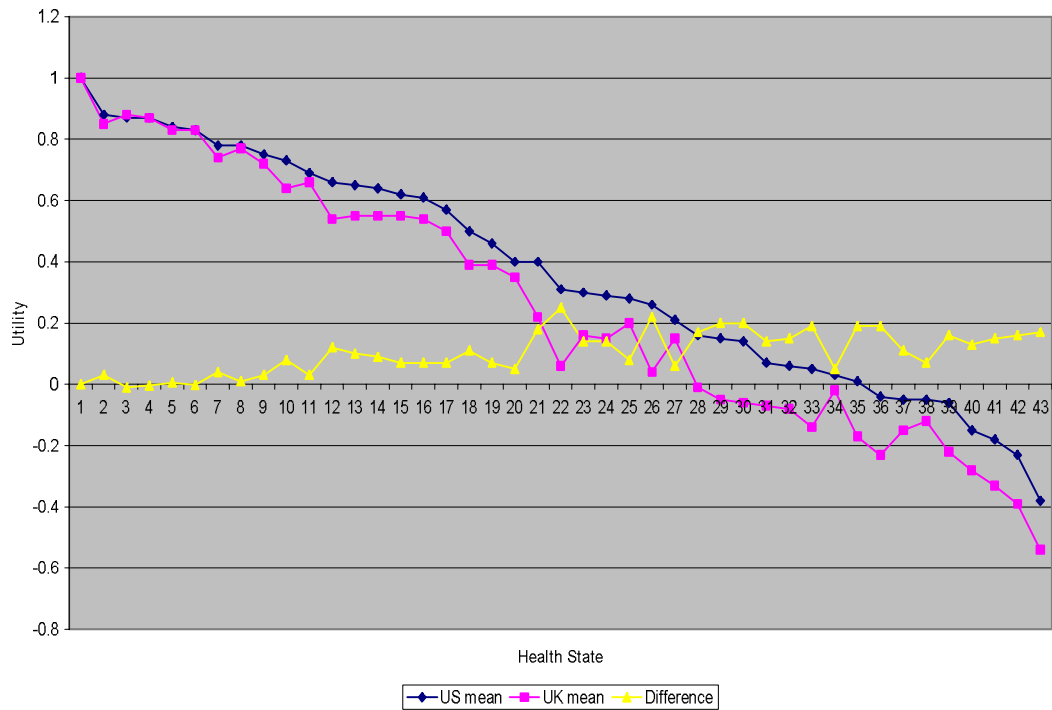

Figure 8. Relationship between the US and UK predicted mean valuations of the 43
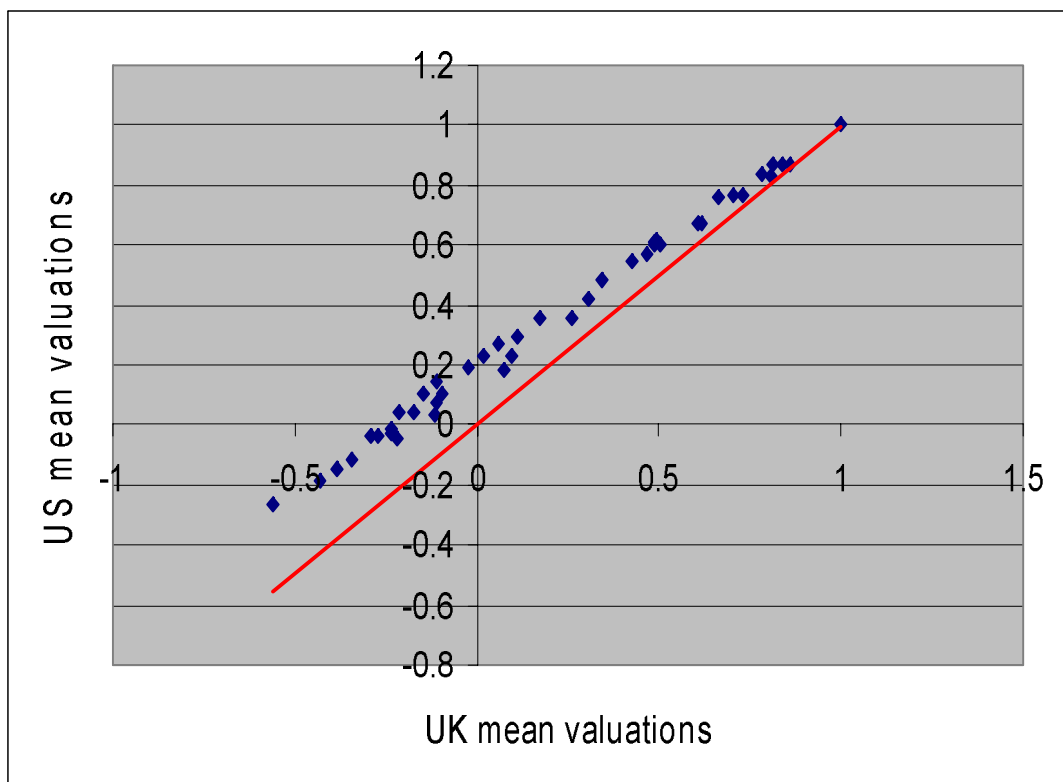EQ-5D health states for the Bayesian model

Figure 9. Relationship between the US and UK predicted mean valuations of the 43 EQ-5D health states for the Johnson et al model
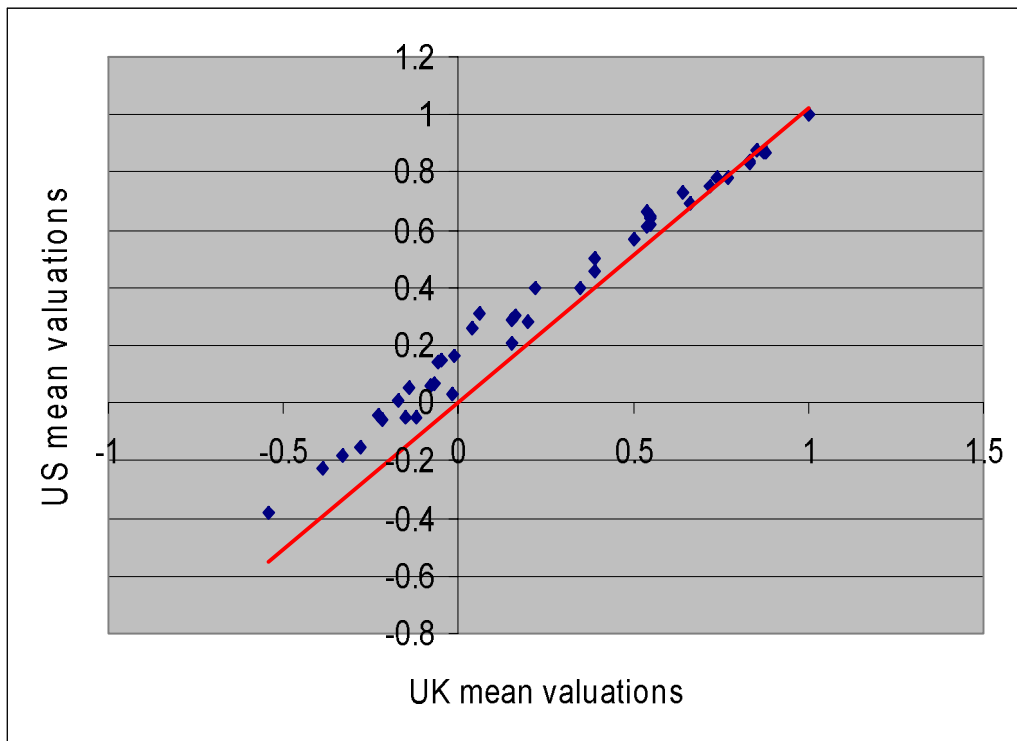

Figure 10. Standardised residuals for each of the 81240 individual health state valuation
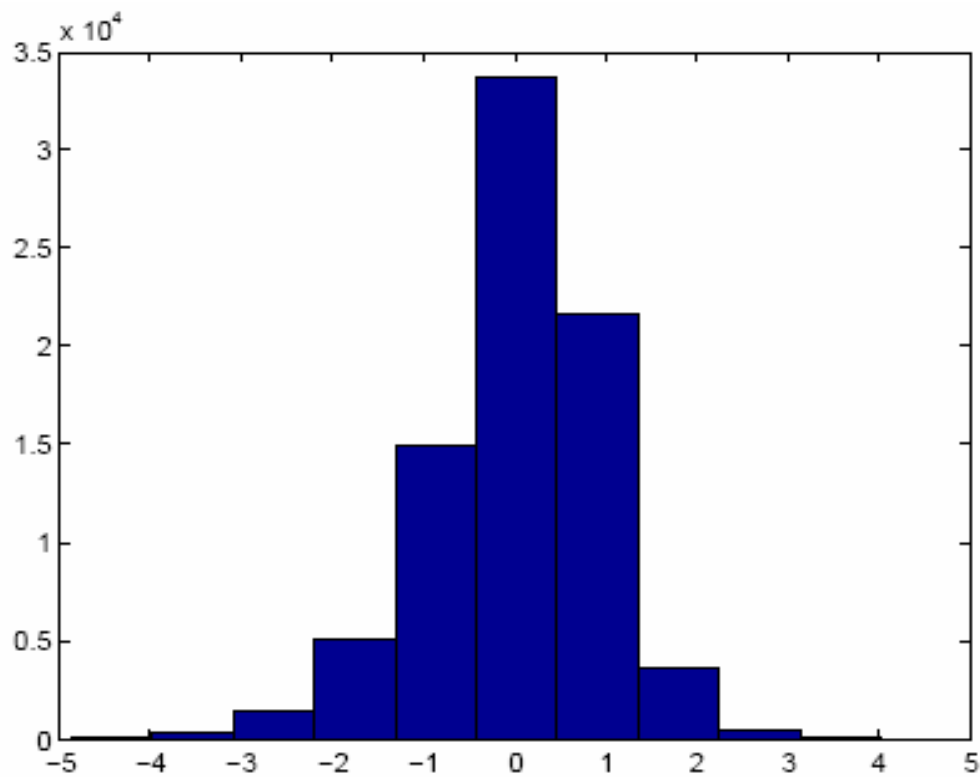
Table 2: Johnson et al (2005) random effect model of US and UK EQ-5D valuations.

	Main Effects Only		Main effects plus interaction terms	
	Coefficient	P value	Coefficient	P value
Intercept	0.576		0.645	
US	0.095		0.011	0.854
Age	0.011		0.012	
Age2	-0.0001		-0.0001	
Sex	0.028	0.022	0.057	
M2	-0.052		-0.073	
M3	-0.342		-0.333	
S2	-0.07		-0.084	
S3	-0.232		-0.229	
U2	-0.09		-0.122	
U3	-190		-0.228	
P2	-0.056		-0.093	
P3	-0.373		-0.454	
A2	-0.055		-0.078	
A3	-0.258		-0.322	
US*Age			-0.001	0.707
US*Age2			<0.0001	0.697
US*Sex			-0.035	0.062
US*M2			0.026	0.0003
US*M3			-0.009	0.437
US*S2			0.016	0.019
US*S3			-0.003	0.752
US*U2			0.04	
US*U3			0.047	
US*P2			0.044	
US*P3			0.097	
US*A2			0.028	0.0002
US*A3			0.077	

Table 3: Prediction of reserved data means

Missing state	True Sample mean	Posterior Mean	Posterior S.D.
12121	0.7641	0.7139	0.0902
22331	0.1109	0.065	0.1146
33232	-0.2228	-0.2028	0.1019
R.M.S.E.	0.040		