



The  
University  
Of  
Sheffield.

School of  
Health  
And  
Related  
Research

## Health Economics & Decision Science (HEDS) Discussion Paper Series

Development of methods for the mapping of utilities using mixture models: An application to asthma

Authors: Laura A Gray, Mónica Hernández Alava, Allan J Wailoo

Corresponding author: Mónica Hernández Alava

ScHARR, University of Sheffield,  
Regent Court, 30 Regent Street, Sheffield, S1 4DA  
Tel: (+44) (0)114 222 0736  
Fax: (+44) (0)114 222 0710  
Email: [monica.hernandez@sheffield.ac.uk](mailto:monica.hernandez@sheffield.ac.uk)

No. 17.03

**Disclaimer:**

This series is intended to promote discussion and to provide information about work in progress. The views expressed in this series are those of the authors. Comments are welcome, and should be sent to the corresponding author.

This paper is also hosted on the White Rose Repository: <http://eprints.whiterose.ac.uk/>

# **Development of methods for the mapping of utilities using mixture models: An application to asthma**

Laura A Gray, Mónica Hernández Alava, Allan J Wailoo

Corresponding author: Mónica Hernández Alava

ScHARR, University of Sheffield,

Regent Court, 30 Regent Street, Sheffield, S1 4DA

Tel: (+44) (0)114 222 0736

Email: [monica.hernandez@sheffield.ac.uk](mailto:monica.hernandez@sheffield.ac.uk)

## ABSTRACT

**Objectives:** To develop methods for mapping to preference-based measures using mixture model approaches. These methods are compared to map from the Asthma Quality of Life Questionnaire (AQLQ) to both EQ5D-5L and HUI-3 as the target health utility measures in an international dataset.

**Methods:** Data from 856 patients with asthma collected as part of the Multi-Instrument Comparison (MIC) international project were used. Adjusted limited dependent variable mixture models (ALDVMMs) and beta-regression based mixture models were estimated. Optional inclusion of the gap between full health and the next value, and a mass point at the next feasible value were explored. Response-mapping could not be implemented due to missing data.

### **Results:**

In all cases, model specifications which formally modelled the gap between full health and the next value were an improvement on those which did not. Mapping to HUI3 required more components in the mixture models than mapping to EQ5D-5L due to its uneven distribution. The optimal beta-based mixture models mapping to HUI3 included a probability mass at the utility value adjacent to full health. This is not the case when estimating EQ5D-5L, due to the low proportion of observations at this point.

### **Conclusion:**

The beta-based mixture models marginally outperformed ALDVMM in this dataset when comparing models with the same number of components. This is at the expense of requiring a larger number of parameters and estimation time. Both model types are able to closely fit the data without biased characteristic of many mapping approaches. Skilled judgment is critical in determining the optimal model. Caution is required in ensuring a truly global maximum likelihood has been identified.

**Keywords:** mapping, EQ5D, HUI, utility

**Acknowledgements:** We are grateful to Jeff Richardson and colleagues for providing data from the Multi-Instrument Comparison Project. This study was supported by the Medical Research Council under grant MR/L022575/1.

## INTRODUCTION

Preference-based outcome measures (PBMs) that allow the calculation of health state utilities are not always administered in studies of clinical effectiveness. However, these outcomes are often preferred by decision makers in order to estimate quality-adjusted life-years (QALYs) for use in cost-effectiveness analysis. 'Mapping', or 'cross-walking', is commonly used to estimate health state utilities when clinical studies have not included any preference based measure (1).

There are two broad approaches to mapping. The direct approach models the health state utility values themselves. The indirect approach, also referred to as response-mapping, models each dimension of the PBM and calculates the predicted utilities as a second, separate step. Response-mapping models require observations (preferably a sizeable number) at all levels of each dimension and this can be a problem for small datasets if there are many different levels in each dimension.

Health state utility values are characterised by unusual distributions; they are commonly skewed, multi-modal, often have a large number of observations at 1 (indicating full health) and a gap between full health and the next feasible value. By definition, they are limited between the range of best and worst health states. Basic regression models are unable to capture all these features which leads to biased estimates of health gain.

Beta regressions can provide flexibility when modelling skewed, bounded PBMs. Basu & Manca proposed the use of single and two-part beta regressions to model PBMs and QALYs(2). The standard beta regression assumes that the dependent variable is only defined in the open interval (0,1) but many PBMs display negative values. Some studies have suggested that a beta regression is inappropriate in these cases(3). Other studies have attempted to overcome this problem by converting *ad hoc* all negative values to zero(4,5) not only ignoring that some health states are worse than death but potentially distorting the distribution due to the well-known sensitivity of beta regressions to pile-ups of values at the boundaries. However, there is a standard transformation<sup>1</sup> in the literature which allows the transformation of values in any open interval into a (0,1) interval(6). After estimation, the expected value is then transformed back to its original scale to obtain the correct predictions. In the area of mapping, this is the approach followed by Kent *et al.* and Khan *et al.*(7,8). Beta-based regression models have been found to be more robust and outperform linear regressions(2,8,9). One significant issue when using beta regressions is how to deal with observations on the boundaries of the feasible utility range. Different methods have been proposed and it is recommended that the sensitivity of the estimates to the different methods is assessed(6). Even though beta regressions can deal with the bounded nature of utility data and can reproduce a variety of shapes, multimodality is difficult to capture.

Mixture models are increasingly being used in the context of mapping because of their flexibility and the ability to capture multimodality(2–6,10,11). Mixtures of normal distributions have been used to model different PBMs such as HUI3(12), EQ-5D-3L(7,9) and SF-6D(9). Some mixture models have been specifically designed for utility mapping such as the adjusted limited dependent variable mixture model (ALDVMM)(10,11,13,14). This employs a mixture of adjusted normal distributions to account for the multimodality of PBMs and includes a number of other useful characteristics. It contains built-in features which account for the peak of observations at full health and the option of a gap in the distribution below that peak. Other mixture models used for mapping include a mixture of Tobit models censored to account for the bounded nature of PBMs with an additional degenerate distribution at perfect health(15). One additional study(8) claims to estimate a limited dependent variable model. However, the model described is not a finite mixture model but a two-part model with an ad-hoc assumption of a normal distribution for values of the dependent variable below 0.3 and a

---

<sup>1</sup>  $y' = (y - a)/(b - a)$  for the interval [a,b].

beta binomial for values at or above. The split at 0.3 is justified via visual inspection of a kernel density plot of the dependent variable. Recently, beta mixture models have also been used in utility mapping with success(7). In general, mixture models have been found to outperform non-mixture models(13–15). One study found some evidence to suggest that beta regression can outperform mixture models which might be in part related to the distributional shape of the health utility measure being used(9).

This study develops knowledge about mapping methods by comparing approaches for estimating two PBMs, EQ5D-5L and HUI3, from the Asthma Quality of Life Questionnaire Sydney (AQLQ-S) score, a clinical asthma measure using data from an international sample(16). Two different classes of mixture models are used: the ALDVMM and extensions to the beta mixture model in Kent *et al* which a) account for the gap in the PBM distributions between full health and the next feasible value and b) allow alternative approaches to deal with observations on the boundary of the beta distribution(7).

All models are estimated using user-written code in Stata via the commands ‘aldvmm’(13) and ‘betamix’(17).

## METHODS

### *Data*

We used data from the Multi Instrument Comparison (MIC) project dataset which includes data on 7,933 observations across 6 countries: Australia, Canada, Germany, Norway, the UK and the USA(16). The data include information on wellbeing, health state utilities and demographics. In addition, respondents who self-reported having specific conditions were asked to answer disease specific questionnaires. In total, 856 respondents self-reported asthma and completed the AQLQ-S questionnaire. Data were available for respondents’ age and sex as well as their EQ5D-5L and HUI3 scores. After removing observations with missing values in any of the required variables, the final sample for analysis was 852.

### *Preference Based Measures*

Both EQ5D-5L and HUI3 are PBMs with health state utilities estimates for each feasible response to their descriptive system. EQ5D-5L covers the same 5 dimensions as the original 3-level version (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) but each dimension has 5 response levels (no problems, slight, moderate, severe, extreme/unable to do). It is designed for self-completion, has a low response burden and is applicable to a range of diseases and treatments. HUI3 is also a self-completion questionnaire with 8 dimensions (vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain). The levels for each dimension vary between 5 and 6. We use the value sets in Devlin *et al*(18) and Furlong *et al*(19) to attach utility values to each health state in EQ5D-5L and HUI3, respectively. For both instruments, a value of 1 represents full health, a value of 0 is considered equivalent to being dead and their values can be negative, representing a state worse than death. Both instruments have a gap between full health and the next feasible health state (0.951 in EQ5D-5L and 0.97258 in HUI3). We refer to this value as the truncation point. The lower limits are -0.281 and -0.36 respectively for EQ5D-5L and HUI3.

### *AQLQ*

The Sydney Asthma Quality of Life Questionnaire (AQLQ-S) was designed as a measure of quality of life for adult patients with asthma. The questionnaire contains 20 questions within 4 domains (symptoms, activity limitation, emotional function and environmental stimuli). Each of the questions allows a response on a 0-4 scale, with 0 representing no problems at all. The scores for each question

are averaged to produce an overall AQLQ-S score between 0 and 4. Although there are many different versions of the AQLQ, the AQLQ-S is recommended by the European Medicines Agency (EMA)(20) and has been validated(21). However, because the scoring is not preference-based, it is not suitable for use in cost-utility analysis.

### Statistical Methods

It was not feasible to conduct response-mapping since this requires observations in each response category of the different dimensions covered by the target descriptive system. In the case of EQ-5D, there are no observations with the worst possible response for self-care. For HUI, there were no observations with the worst possible response for vision.

We compared two approaches to direct mapping, both based on mixture models. The first is an adjusted limited dependent variable mixture model (ALDVMM) implemented using the publicly available Stata command ‘aldvmm’(13). This model has previously been applied for mapping across a range of clinical areas, including rheumatoid arthritis(13,14), osteoarthritis(22), ankylosing spondylitis(23) and traumatic brain injury(24). It has been shown to outperform other methods (linear regression, Tobit and response-mapping). The ALDVMM is a bespoke model developed specifically for utility mapping and the Stata function includes a number of user-specified options to tailor the method to the target utility instrument and country specific tariff of interest. This includes specifying the next feasible value after full health(13), the “truncation point”, thus creating the typical gap seen in PBMs. There is the option to specify no truncation and therefore allow each component of the mixture model to be fully continuous up to the highest feasible value of 1 for full health. The method has previously been described in detail(14). In brief, ALDVMM is a mixture of adjusted, normal distributions for use when the dependent variable is limited above at 1 (full health) and below, in this case at -0.205 for EQ5D-5L and -0.36 for HUI3. As well as estimating the model with different numbers of components, we also estimated it with and without truncation.

The second model we used is a beta-based mixture model estimated via the user-written Stata command ‘betamix’(17) which is a generalisation of the truncated inflated beta regression model introduced in Pereira et al(25). This is a two-part model consisting of a multinomial logit model and a beta mixture model. A beta distribution cannot deal with observations at the boundaries because the log-likelihood is undefined at these points. The addition of the multinomial logit model to the beta mixture allows for these observations and a mass of observations at full health. The model assumes a limited dependent variable  $y_i$  for each individual  $i$  defined at point 1 and the interval  $[a, \tau]$  where  $a < \tau < 1$  and can be written as

$$g(y_i|x_{i1}, x_{i2}, x_{i3}) = \begin{cases} P(y_i = a|x_{i3}), & y_i = a \\ P(y_i = \tau|x_{i3}), & y_i = \tau \\ P(y_i = 1|x_{i3}), & y_i = 1 \\ \left[1 - \sum_{s=a,b,\tau} P(y_i = s|x_{i3})\right] h(y_i|x_{i1}, x_{i2}), & y_i \in (a, \tau) \end{cases} \quad (1)$$

with probabilities

$$P(y_i = k|x_{i3}) = \frac{\exp(x'_{i3}\gamma_k)}{1 + \sum_{s=a,p,b} \exp(x'_{i3}\gamma_s)} \quad (2)$$

where  $x_{i3}$  is a vector of variables influencing the probabilities,  $\gamma_k$  is a vector of coefficients and  $s$  refers to each section of the distribution. For identification, the coefficients corresponding to the continuous part of the distribution are set to zero. The probability density function for the continuous part of the distribution has probability density function  $h(\cdot)$  made up of a mixture of C-components

each representing a beta distribution, with mean  $\mu_{c_i}$  and precision parameter  $\phi_c$ ,  $c = 1, \dots, C$ , such that

$$h(y_i | x_{i1}, x_{i2}) = \sum_{c=1}^C (P(c | x_{i2}) f(y_i | x_{i1} \beta_c, \phi_c, a, \tau)) \quad (3)$$

where  $f(\cdot)$  is a beta density with alternative parameterisation and  $C$  is the number of components included in the analysis. Component membership is determined using a second multinomial logit model, such that

$$P(c | x_{i2}) = \frac{\exp(x'_{i2} \delta_c)}{\sum_{j=1}^C \exp(x'_{i2} \delta_j)} \quad (4)$$

where  $x_{i2}$  is a vector of variables influencing the probability of component membership and  $\delta_c$  is a vector of corresponding coefficients. Again, one set of coefficients is set to zero for identification.

The model is not constrained to the (0,1) range but transforms values to the relevant interval for the target utility instrument (-0.281 to 1 for EQ5D-5L and -0.36 to 1 for HUI). It is capable of producing estimates at either of the feasible limits, though for health utilities this is most relevant for mass points at full health (1). As with the ALDVMM, betamix also allows the specification of a gap between full health and the next feasible health state and for a mass of observations at this truncation point. Although it is possible to include a probability mass at the lower utility limit, for both model types, we did not include this here because our sample contained no observations with values at the lower utility limit for either PBM.

We estimated different specifications of each model type, with different numbers of components and with and without probability masses at certain points of the distribution. We included AQLQ-S summary score, age, age-squared and sex as covariates in all parts of the model. For each of the models we are able to use simulated data to create figures showing the conditional distribution functions for each of the models.

Preferred models were selected using a variety of fit statistics: Akaike and Bayesian Information Criteria (AIC and BIC), root mean squared error (RMSE), mean absolute error (MAE) and mean error (ME). We assessed fit across the distribution of disease severity. We compare the conditional distribution function of the observed data with the one derived from the estimated model. This builds on previous work in the area which focus on the summary measures(5). In many cases these criteria each support different models and so judgement must be used in determining the preferred model.

## RESULTS

The final sample consisted of 852 observations (see Table 1) of which 62.3% were female. Age ranged from 18 to 89 years. Whilst the AQLQ spanned the entire range of feasible values (0-4), neither EQ5D-5L or HUI3 did.

**Figure 1: Distribution of a) EQ5D-5L and b) HUI3**

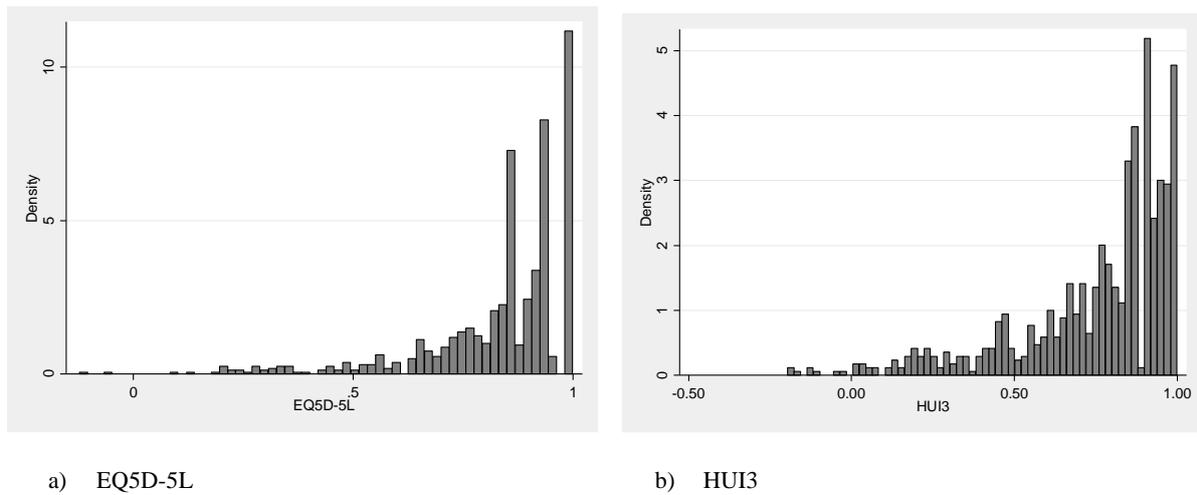


Figure 1a and 1b show the distributions of EQ5D-5L and HUI3, respectively. Both HUI3 and EQ5D-5L exhibit mass points at the upper full-health limit: 20.9% in EQ5D-5L and 9.5% in HUI3. For EQ5D-5L, there was no significant mass of observations at the truncation point (0.951). Almost 6% of observations were at the HUI3 truncation point (0.973). There were a relatively large number of observations with an EQ5D-5L at 0.942, 0.924 and 0.866, associated with slight problems with anxiety and depression and/or pain and discomfort.

**Table 1: Sample summary statistics**

	Mean (std. dev.)	Min	Max
<b>AQLQ-S</b>	0.7085 (0.7766)	0	4
<b>EQ5D-5L</b>	0.8425 (0.1693)	-0.073	1
<b>HUI3</b>	0.7560 (0.2408)	-0.1958	1
<b>Age (years)</b>	43.03 (15.00)	18	89
<b>Country</b>	Number of observations (%)		
<b>1</b>	141 (16.55)	-	-
<b>2</b>	150 (17.61)	-	-
<b>3</b>	150 (17.61)	-	-
<b>4</b>	138 (16.20)	-	-
<b>5</b>	126 (14.79)	-	-
<b>6</b>	147 (17.25)	-	-

AQLQ-S: Asthma quality of life questionnaire-Sydney, EQ5D-5L: EuroQol 5-dimension 5-level questionnaire, HUI3: Health utilities index mark 3.

**Table 2: Model specifications and model choice criteria (n=852)**

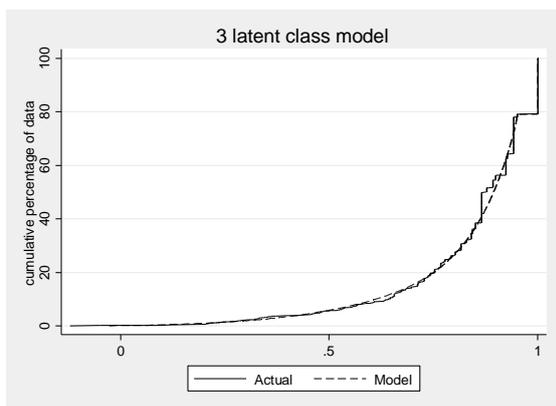
# component s	specification	Log-likelihood	# param	RMSE	MAE	ME	AIC	BIC
EQ5D-5L Betamix								
3	Probability mass at full health	527.68	33	0.1430	0.1001	-0.0003	-989.36	-832.69
3	Probability mass at full health and truncation point	436.00	38	0.1425	0.1003	0.0005	-796.00	-615.59
4	Probability mass at full health	538.87	44	0.1429	0.1002	0.0003	-989.75	-780.86
4 <sup>‡</sup>	Probability mass at full health and truncation point	401.81	45	0.1474	0.1018	-0.0012	-713.62	-499.98
EQ5D-5L ALDVMM								
3	Bounded	322.42	28	0.1439	0.1006	0.0003	-588.84	-455.90
4	Bounded	336.64	39	0.1439	0.1004	$\frac{0.0000}{3}$	-595.28	-410.13
HUI3 Betamix								
3	Probability mass at full health	708.05	33	0.2081	0.1566	0.0018	-1350.09	-1193.42
3	Probability mass at full health and truncation point	207.64	38	0.2081	0.1563	0.0024	-339.28	-158.88
4	Probability mass at full health	727.92	44	0.2076	0.1562	0.0017	-1367.80	-1158.95
4	Probability mass at full health and truncation point	224.48	49	0.2067	0.1548	0.0009	-350.96	-118.33
HUI3 ALDVMM								
3	Age included	192.51	28	0.2076	0.1556	$\frac{0.0004}{1}$	-329.03	-196.10
4	Age included	212.42	39	0.2071	0.1550	$\frac{0.0003}{3}$	-346.85	-161.69
3	No age in probability variables $x_2$	189.87	24	0.2082	0.1563	$\frac{0.0005}{4}$	-331.75	-217.80
4	No age in probability variables $x_2$	201.09	33	0.2082	0.1563	$\frac{0.0002}{6}$	-336.18	-179.51

RMSE = Root Mean Squared Error, MAE = Mean Absolute Error, ME = Mean Error, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion. All models outlined here include a truncation at the best possible health state other than full health. <sup>‡</sup>This model would not converge with the AQLQ score in the probabilities parameters. The results presented here are for a model without AQLQ in the probabilities.

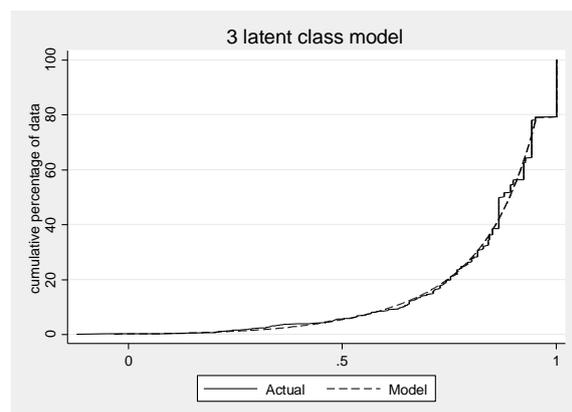
*EQ5D-5L*

We found that models which formally included the gap between full health and the next feasible value outperformed those that did not, using both ALDVMM and beta-base mixture models. We therefore concentrate on comparing alternative specifications of models that included this gap. Beta mixture models required a specified probability mass at full health in order to ensure they estimated the correct proportion of observations at full health. Table 2 shows model performance criteria for 3 and 4 component models, each with inclusion of truncation, a mass point at full health and with and without a further probability mass at the truncation point (0.951). Differences in measures of “error” between 3 and 4 component models were small, with BIC lower for the 3 component model(14). The model which does not include a probability mass at the truncation point appears to better predict the lower end of the EQ5D distribution. This can be seen in the conditional distribution function graphs in Figure 2a and the plots of mean predicted versus observed fit in Figure 3a. This is because there are a relatively small number observations at the truncation point but a large proportion of observations at the value just below the truncation point (13.73% at 0.944). If this spike in observations were at the truncation point itself, the model which included a probability mass at the truncation point might have shown better fit. For these reasons, the optimal beta-based model has 3 components and a probability mass at full health but not at the truncation point.

**Figure 2a: Conditional Distribution Functions for models Estimating EQ5D-5L with betamix (observed vs simulated data (1,000 observations))**

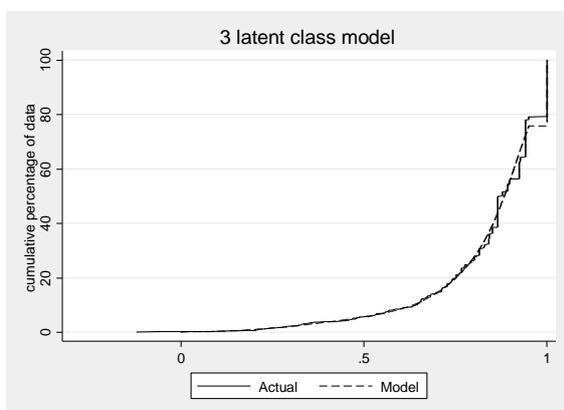


3 component betamix model for EQ5D-5L with truncation and probability mass at full health

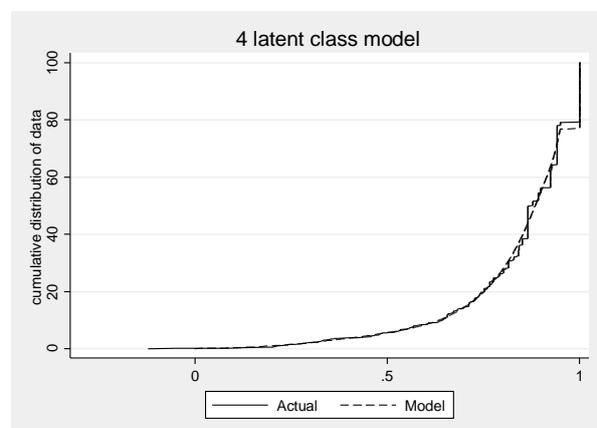


3 component betamix model for EQ5D-5L with truncation and probability mass at full health and truncation point

**Figure 2b: Conditional Distribution Functions for models Estimating EQ5D-5L with ALDVMMs (observed vs simulated data (1,000 observations))**

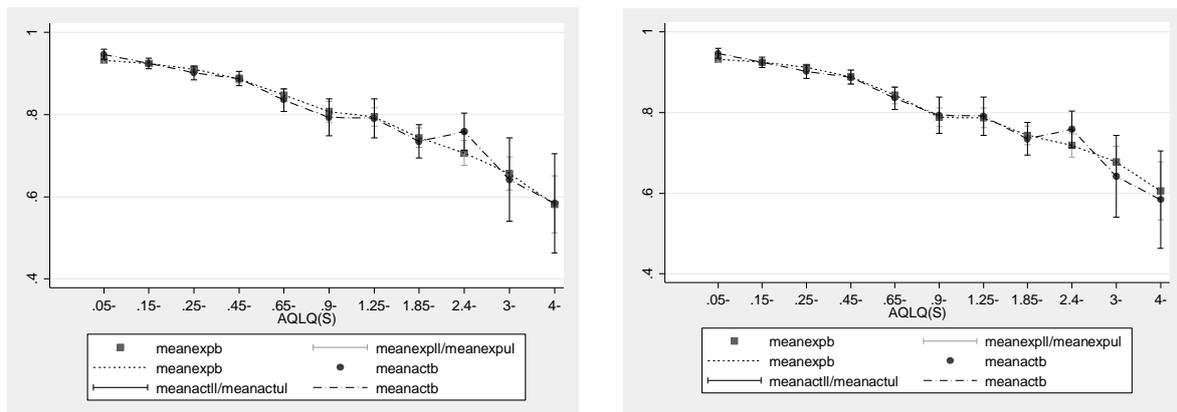


3 component ALDVMM for EQ5D-5L with truncation



4 component ALDVMM for EQ5D-5L with truncation

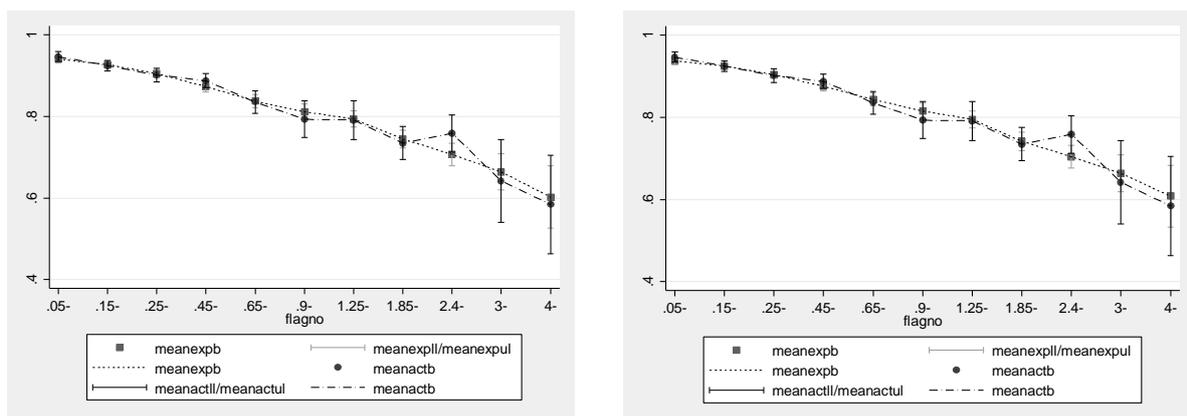
**Figure 3a: Mean Fit vs Mean Observed PBM by AQLQ for models Estimating EQ5D-5L with betamix**



3 component betamix model for EQ5D-5L with truncation and probability mass at full health

3 component betamix model for EQ5D-5L with truncation and probability mass at full health and truncation point

**Figure 3b: Mean Fit vs Mean Observed PBM by AQLQ for models Estimating EQ5D-5L with ALDVMM**



3 component ALDVMM for EQ5D-5L with truncation

4 component ALDVMM for EQ5D-5L with truncation

Results for 3 and 4 component ALDVMMs are shown in Table 2. The 4 class model offers improvements in RMSE, MAE and ME but has a higher BIC. Figures 2b and 3b show that both models fit the data closely, suggesting the 3 component model is preferred. Figure 2b shows some disparity between the distribution from the model and the data at the upper end of EQ5D-5L. This occurs not at the full health value (the data has 21% of observations here compared to 24% in the simulated data in the models) but at values just below the truncation point.

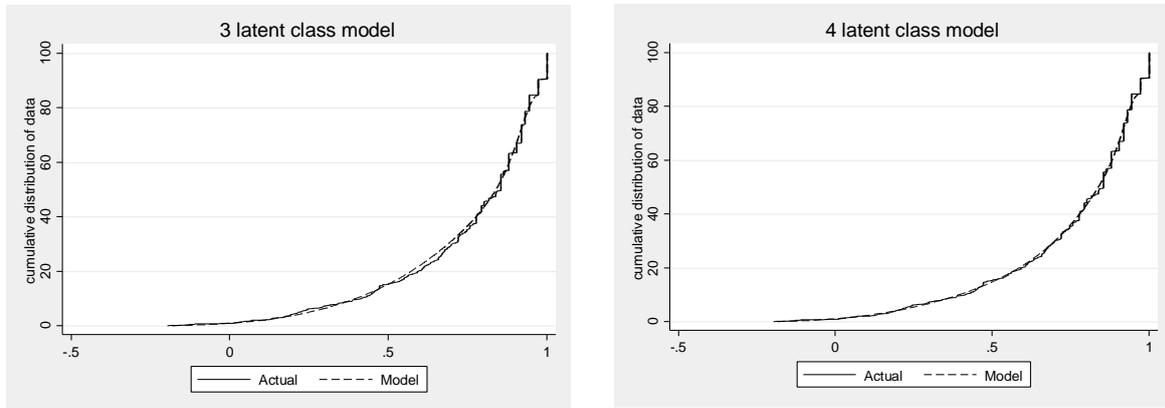
*HUI3*

As with EQ5D-5L, we found that models which formally included the gap between full health and the next feasible value were preferred to those which did not, having a much lower AIC and BIC. This applied to the beta-based mixture model and ALDVMM.

Beta-based models without a probability mass at full health had difficulty fitting the correct number of observations at this value and so we only report models which explicitly modelled this gap and include a probability mass at full health. The model which consistently produced the smallest errors was the 4 component model with probability masses at full health and at the truncation point (0.973). However, although AIC suggested that the fourth component is beneficial, BIC suggests that 3

components is preferred. The simulated graphs in Figure 4a and the plots of the means in Figure 5a show a clear improvement in the model with the additional fourth component, particularly towards the lower end of the HUI3 distribution. We consider the 4 component model to be the optimal beta mixture model.

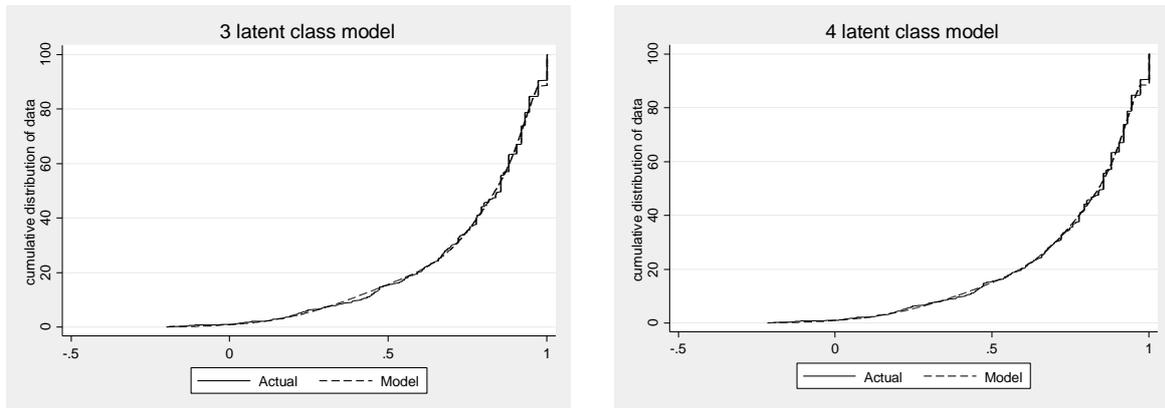
**Figure 4a: Conditional Distribution Functions for models Estimating HUI3 with betamix (observed vs simulated data (1,000 observations))**



3 component betamix model for HUI3 with truncation and probability mass at full health only and truncation point

4 component betamix model for HUI3 with truncation and probability mass at full health and truncation point

**Figure 4b: Conditional Distribution Functions for models Estimating HUI3 with ALDVMM (observed vs simulated data (1,000 observations))**



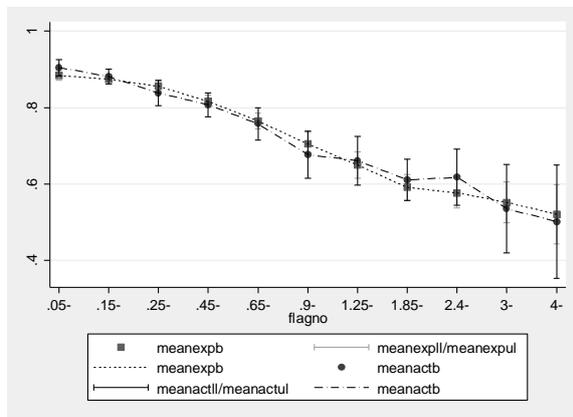
3 component ALDVMM for HUI3 with truncation

4 component ALDVMM for HUI3 with truncation

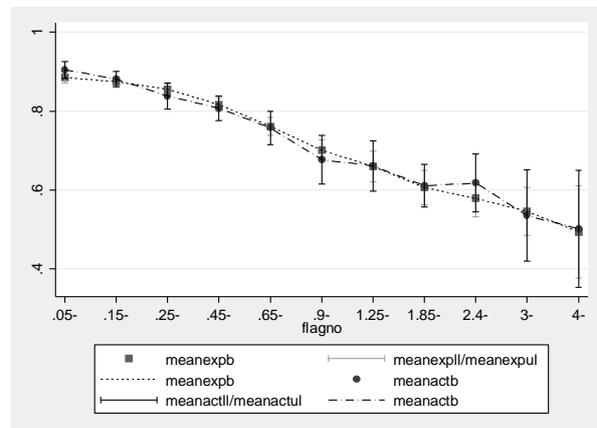
As with the beta mixture models, 4 component ALDVMMs produced lower errors than 3 component equivalents. In all ALDVMMs estimating HUI3, the coefficients for the variables influencing component membership were statistically insignificant: for example age had p-values all in excess of 0.03. We therefore investigated the use of different variables to predict component membership. Results are displayed in Table 2 for models which do not include age to predict component probabilities as well as those which do. Whilst AIC and BIC generally favour the exclusion of age, other measures of error are worse. Figures 4b and 5b show a marked difference between these models, and suggest that age should remain an explanatory variable for the probabilities because they considerably improve the fit of the model. The evidence indicates that the observed statistical insignificance associated with age may be related to the limited sample size.

All models have the expected sign and produce simulated data which is a good prediction of the sample data. The estimated coefficients for the 4 preferred models are displayed in the web appendix along with a Stata .do file which allows users to enter their own data to predict EQ5D-5L and HUI3 using these preferred models.

**Figure 5a: Mean Fit vs Mean Observed PBM by AQLQ for models Estimating HUI3 using betamix**

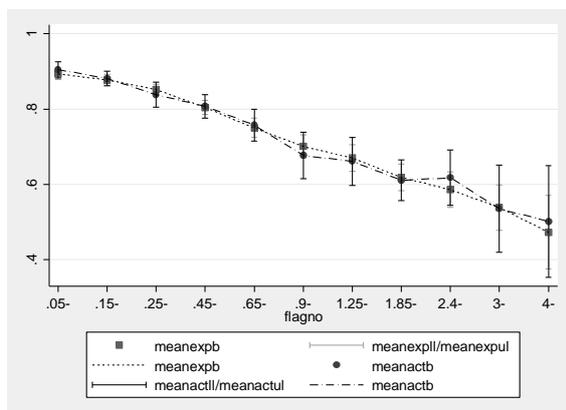


3 component betamix model for HUI3 with truncation and probability mass at full health and truncation point

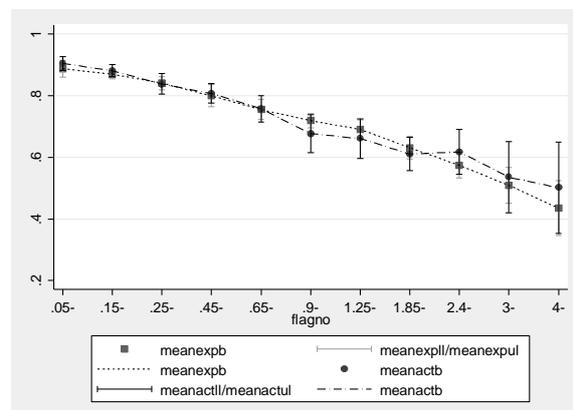


4 component betamix model for HUI3 with truncation and probability mass at full health and truncation point

**Figure 5b: Mean Fit vs Mean Observed PBM by AQLQ for models Estimating HUI3 using ALDVMM**



4 component ALDVMM for HUI3 with truncation



4 component ALDVMM for HUI3 with truncation (age removed from probability variables  $x_2$ )

## DISCUSSION

We compared different mapping methods using data collected from people with asthma. Both EQ5D-5L and HUI3 were considered as target utility instruments, from the AQLQ-S. We focussed on direct mapping methods because response-mapping was not feasible in this sample, for either instrument. A move to EQ5D-5L, compared to EQ5D-3L, is likely to reduce the feasibility of applying response-mapping in future, since data samples will more often fail to span all the levels described in the more detailed descriptive system.

Beta-based mixture models and ALDVMMs were estimated. Within these classes of model-type we compared different numbers of components in the mixture models with and without a specified gap between full health and the next feasible value. For the beta based models, where the gap was

specified, we also considered whether there needed to be a specified probability mass at the truncation point.

Models with truncation to create a gap between full health and the next feasible value were universally preferred. This gap is part of the “bespoke” nature of both the ALDVMM and the beta-based mixture model and, whilst the importance of the method has been demonstrated in several studies using EQ5D-3L(14), the gap is more pronounced in that instrument (0.12). This contrasts with 0.049 and 0.027 in the EQ5D-5L and the HUI3 respectively. The next largest gap between feasible values is 0.024 in EQ5D-5L and 0.027 in HUI3. To our knowledge no mapping study which uses a mixture of beta distributions has accounted for the gap between full health and the next feasible utility value. We find that it remains important for mapping models to explicitly reflect this characteristic of utility instruments, even when the gap is relatively small. Others have claimed the gap is sufficiently small not to warrant formal inclusion in the model(5). Furthermore, the finding is relevant to all mapping methods and not simply those that utilise mixture model approaches.

Beta regression mixture models were identified that better fitted both EQ5D-5L and HUI3 than ALDVMMs. Preferred models allowed an inflated number of observations at the next feasible value below full health. This feature adds further flexibility to the models but caution needs to be exercised in interpreting this finding. Theoretically, the addition of a mass point to the beta model is similar to adding an additional component to the ALDVMM. In this sense, the beta mixture approach is more artificial than the ALDVMM and comparisons between models with the same number of components are not necessarily comparisons of like-with-like. The addition of mass points at selected points in the distribution does offer a means to improve fit but this requires the addition of more parameters compared and is also less generalizable: there is a risk of over fitting to the data. In general we would recommend caution in the inclusion of probability masses at ANY point in the distribution without some theoretical rather than empirical justification for doing so: by ‘theoretical justification’ we mean that one is clear why such peaks may be relevant, for example the nature of the disease leads to bunching at a certain health state utility value. The increased number of parameters required by the beta-based mixture model means that for smaller datasets there is a danger that it might be more difficult to identify. For example, we attempted to estimate EQ5D-5L using a four component model with probability masses at full health and the truncation point, but this model would not converge when AQLQ-S was included in the component membership probabilities. To reduce the number of parameters, we estimated a model without AQLQ-S in the probabilities and achieved a much worse fit as a result.

Our results show the importance of considering the distribution when choosing the most appropriate model. The proportion of observations at the truncation point in a sample should be considered when choosing a model, in particular when using the beta-based mixture model. We found that where there were lots of observations at the truncation point (as with HUI3) then a probability mass needed including at the truncation point. The inclusion of this probability mass did not improve the model for EQ5D-5L which had very few observations at the truncation point. However, as discussed above, theoretical justification is needed to determine whether these probability masses are generalizable. We do not know if this is generalizable to all applications of HUI3 and EQ5D-5L in asthma patients or specific to this sample. Further research is needed to determine the extent to which these results are generalizable.

Other studies have been conducted using asthma outcomes. Tsuchiya et al estimated EQ5D-3L from the less commonly reported 32-item AQLQ-McMaster score(26). As well as direct mapping using a linear regression, this study also carried out the first response-mapping we are aware of. Our results are not directly comparable with this study because the AQLQ scores differ. We are able to compare our results to Kaambwa et al who used the same data as in our study. Kaambwa et al. mapped AQLQ-S onto both EQ5D-5L and HUI3, among other health state utilities(5) using data from the MIC. The

mixture models reported here outperformed every equivalent model reported by Kaambwa et al; that is all models which used the AQLQ-S summary score as an explanatory variable, using the measures of model fit describe earlier. They used four simple methods: ordinary least squares (OLS), censored least absolute deviations (CLAD), generalised linear model (GLM) and a beta binomial (BB) regression model.

Mixture models are much more flexible than typically used mapping models. The methods we test constrain model outputs to the feasible range between full health and the worst health state, with the ability to have large masses at the extremes. Modelling the gap between full health and the next feasible health state adds additional flexibility and further restricts outputs to the feasible range for the utility instrument. When using these more complex models it is important to consider the characteristics of the data and to search for the most appropriate model, both through using different specifications of the models, but also ensuring that the global maximum likelihood is found(27). In our search for a global maximum, we found other maxima which included mass points at the top of the distribution. This predicted the number of observations in full health very well but the overall fit was much worse and the graphs for the means showed clear misspecification.

Our results show that each of the chosen models is an improvement on more traditionally used linear predictions. Both types of mixture model used in this study are able to closely fit the data without the biased performance characteristic of many commonly employed mapping approaches. The beta-based mixture models outperformed the ALDVMM models but at the expense of increasing the number of parameters as well as estimation time. Skilled judgment is critical in determining the optimal model. Caution is required in ensuring a truly global maximum likelihood has been identified.

## References

1. Kearns B, Ara R, Wailoo A, Manca A, Hernández Alava M, Abrams K, et al. Good practice guidelines for the use of statistical regression models in economic evaluations. *Pharmacoeconomics*. 2013;31(8):643–52.
2. Basu A, Manca A. Regression Estimators for Generic Health-Related Quality of Life and Quality-Adjusted Life Years. *Med Decis Mak* [Internet]. 2012;32(1):56–69. Available from: <http://mdm.sagepub.com/cgi/doi/10.1177/0272989X11416988>
3. Young TA, Mukuria C, Rowen D, Brazier JE, Longworth L. Mapping Functions in Health-Related Quality of Life: Mapping From Two Cancer-Specific Health-Related Quality-of-Life Instruments to EQ-5D-3L. *Med Decis Making* [Internet]. 2015;0272989X15587497 – . Available from: <http://mdm.sagepub.com/content/early/2015/05/20/0272989X15587497.long>
4. Khan I, Morris S. A non-linear beta-binomial regression model for mapping EORTC QLQ-C30 to the EQ-5D-3L in lung cancer patients: a comparison with existing approaches. *Health Qual Life Outcomes* [Internet]. 2014;12:163. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4234877/>
5. Kaambwa B, Chen G, Ratcliffe J, Jezzi A. Mapping Between the Sydney Asthma Quality of Life Questionnaire (AQLQ-S) and Five Multi-Attribute Utility Instruments (MAUIs). *Pharmacoeconomics* [Internet]. Springer International Publishing; 2016; Available from: "http://dx.doi.org/10.1007/s40273-016-0446-4
6. Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol Methods*. 2006;11(1):54–71.
7. Kent S, Gray A, Schlackow I, Jenkinson C, McIntosh E. Mapping from the Parkinson's Disease Questionnaire PDQ-39 to the Generic EuroQol EQ-5D-3L: The Value of Mixture Models . *Med Decis Mak* [Internet]. 2015;(October):902–11. Available from: <http://mdm.sagepub.com/content/early/2015/04/29/0272989X15584921.abstract>
8. Khan I, Morris S, Pashayan N, Matata B, Bashir Z, Maguirre J. Comparing the mapping between EQ-5D-5L, EQ-5D-3L and the EORTC-QLQ-C30 in non-small cell lung cancer patients. *Health Qual Life Outcomes* [Internet]. Health and Quality of Life Outcomes; 2016;14(1):60. Available from: <http://hqlo.biomedcentral.com/articles/10.1186/s12955-016-0455-1>
9. Khan KA, Madan J, Petrou S, Lamb SE. Mapping between the Roland Morris Questionnaire and generic preference-based measures. *Value Heal* [Internet]. Elsevier; 2014;17(6):686–95. Available from: <http://dx.doi.org/10.1016/j.jval.2014.07.001>
10. Hernández Alava M, Wailoo A, Wolfe F, Michaud K. The relationship between EQ-5D, HAQ and pain in patients with rheumatoid arthritis. *Rheumatol (United Kingdom)*. 2013;52(5):944–50.
11. Hernandez Alava M, Wailoo A, Wolfe F, Michaud K. A Comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. *Med Decis Mak* [Internet]. 2014;34(7):919–30. Available from: <http://eprints.whiterose.ac.uk/74543/nhttp://www.ncbi.nlm.nih.gov/pubmed/24025662/nhttp://mdm.sagepub.com/cgi/doi/10.1177/0272989X13500720>
12. Austin P, Escobar M. The use of finite mixture models to estimate the distribution of the health utilities index in the presence of a ceiling effect. *J Appl Stat*. 2003;30(8):909–23.
13. Hernandez Alava M, Wailoo AJ. The Stata Journal. *Stata J* [Internet]. 2015;15(3):737–50. Available from: <http://ideas.repec.org/a/tsj/stataj/v7y2007i4p465-506.html>
14. Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: Adjusted limited

- dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value Heal* [Internet]. Elsevier Inc.; 2012;15(3):550–61. Available from: <http://dx.doi.org/10.1016/j.jval.2011.12.014>
15. Coca Perrailon M, Shih Y-CT, Thisted R a. Predicting the EQ-5D-3L Preference Index from the SF-12 Health Survey in a National US Sample: A Finite Mixture Approach. *Med Decis Mak* [Internet]. 2015;(773):0272989X15577362 – . Available from: <http://mdm.sagepub.com/cgi/doi/10.1177/0272989X15577362>
  16. Richardson J, Iezzi A, Khan M, Maxwell A. Cross-national comparison of twelve quality of life instruments. MIC paper 1: back-ground, questions, instruments. *Researchpaper76*, Melbourne, VIC; 2012.
  17. Gray LA, Hernandez-Alava M. BETAMIX: A command for fitting mixture regression models for bounded dependent variables using the beta distribution. (under submission). 2016;
  18. Devlin N, Shah K, Feng Y, Mulhern B, Hout B van. Valuing health-related quality of life : An EQ-5D-5L value set for England. *Off Heal Econ*. 2016;16(01):1–22.
  19. Furlong W, Feeny D, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Ann Med*. Working Paper No. 98-11; 2001;33(5).
  20. Medicines Agency E. Guideline on the clinical investigation of medicinal products for the treatment of asthma. London, UK; 2015. Report No.: CHMP/EWP/2922/01.
  21. Everhart RS, Smyth JM, Santuzzi AM, Fiese BH. Validation of the Asthma Quality of Life Questionnaire with momentary assessments of symptoms and functional limitations in patient daily life. *Respir Care* [Internet]. 2010;55(4):427–32. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20406510>
  22. Wailoo A, Hernandez Alava M, Escobar Martinez A. Modelling the relationship between the WOMAC osteoarthritis index and EQ-5D. *Health Qual Life Outcomes* [Internet]. 2014;12(1):37. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24621311>
  23. Wailoo A, Hernández M, Philips C, Brophy S, Siebert S. Modeling Health State Utility Values in Ankylosing Spondylitis: Comparisons of Direct and Indirect Methods. *Value Heal* [Internet]. Elsevier; 2015;18(4):425–31. Available from: <http://dx.doi.org/10.1016/j.jval.2015.02.016>
  24. Fuller G, Hernandez M, Pallot D, Lecky F, Stevenson M, Gabbe B. Health state preference weights for the Glasgow Outcome Scale following traumatic brain injury: a systematic review and mapping study. *Value Heal*.
  25. Pereira GHA, Botter DA, Sandoval MC. The Truncated Inflated Beta Distribution. *Commun Stat - Theory Methods* [Internet]. 2012;41(5):907–19. Available from: <http://dx.doi.org/10.1080/03610926.2010.530370>
  26. Tsuchiya A, Brazier J, McColl E, Parkin D. Deriving preference-based single indices from non-preference based condition-specific instruments: Converting AQLQ into EQ5D indices. *Sheff Heal Econ Gr -Discussion Pap Ser*. 2002;02/1.
  27. McLachlan G, Peel D. *Finite Mixture Models*. New York: Wiley; 2005.