Monograph:
Young, T.A., Rowen, D., Norquist, J. et al. (2010) HEDS Discussion Paper 09-15: Developing preference-based health measures: using Rasch analysis to generate health state values. Discussion Paper. Quality of Life Research

HEDS Discussion Paper 09-15

HEDS Discussion Paper 09/15

Published paper
Young TA, Rowen D, Norquist J, Brazier JE. Developing preference-based health measures: using Rasch analysis to generate health state values. Quality of Life Research 2010;19: 907-17.

ScHARR

SCHOOL OF HEALTH AND RELATED RESEARCH

Developing preference-based health measures: using Rasch analysis to generate health state values

Young TA[1,2], Rowen D[1], Norquist J[3], Brazier JE[1*]

1. School of Health and Related Research, University of Sheffield, Sheffield S1 4DA, UK
2. Yorkshire and Humber Research Design Service, University of Sheffield, S1 4DA, UK
3. Merck & Co. Inc., One Merck Drive, Whitehouse Station, New Jersey, USA 08889

* Correspondence to: Professor John Brazier, Health Economics and Decision Science, University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK
Telephone: +44 (0)114 222 0726
Fax: +44 (0)114 272 4095
E-mail: j.e.brazier@sheffield.ac.uk

Abstract

Background/aims: Condition specific measures may not always have independent items, and existing techniques of developing health state values from these measures are inappropriate when items are not independent. This study develops methods for deriving and valuing health states for a preference-based measure.

Methods: Three key stages are presented: Rasch analysis is used to develop a health state classification system and identify a set of health states for valuation. A valuation survey of the health states using time-trade-off (TTO) methods is conducted to elicit health state values. Finally, regression models are applied to map the relationship between mean TTO values and Rasch logit values. The model is then used to estimate health state values for all possible health states. Methods are illustrated using the Flushing Symptoms Questionnaire (FSQ).

Results: Rasch models were fitted to 1270 responders to the FSQ and a series of 16 health states identified for the valuation exercise. An ordinary least squares model best described the relationship between mean TTO values and Rasch logit values. (R2 = 0.958; Root mean square error = 0.042).

Conclusions: We have shown how the valuation of health states can be mapped onto the Rasch scale in order to value all states defined by the FSQ. This should significantly enhance work in this field.

Key words: Rasch analysis; preference-based measures of health; health states; health related quality of life; flushing

* * * * *

Acknowledgements: The authors gratefully acknowledge the help of Dr Vasilisa Sazonov and Dr. Thomas Rhodes from Merck & Co., for their help in completing this study. This study was funded by Merck & Co. Inc and in particular by Global Outcomes Research & Reimbursement department.

Introduction

Preference-based measures (PBM) of health have become a common means of generating health state values for calculating Quality Adjusted Life Years (QALYs). A widely used method to obtaining health state values is to administer one of the generic PBMs of health in a clinical study [1], such as the EQ-5D [2], HUI3 [3] or SF-6D [4]. However, general measures of health have been found to be inappropriate or insensitive for measuring QALYs for some medical conditions [5]. Therefore, many clinicians and researchers prefer condition specific measures. Most condition specific measures are not preference-based and cannot be used to derive the 'quality adjustment weight' for use in QALYs. Thus, there has been increasing interest in developing health state values from condition specific measures.

Previous attempts [6, 7] to develop a preference-based condition specific measure from an existing measure have used Rasch modelling and have adopted procedures originally applied to the SF-36 in the development of the SF-6D preference-based index [4, 8]. This involves the derivation of a multi-dimensional health state classification using a selection of items from the parent instrument. A sample of states defined by this classification are then selected for valuation by a statistical design in order to permit the estimation of a model for predicting the values of all states defined by the health state classification. This approach of sampling selected health states assumes the dimensions of the health state classification are independent, otherwise many of the states generated by the statistical design will be non-sensible (e.g. including combinations such as feeling low most of the time ***and*** feeling full of life all of the time).

This paper presents a study that develops a health state classification from an existing measure of flushing symptoms that is unidimensional, consisting of health states that are dependent of each other and thus particularly prone to this problem. It uses a combination of Rasch analysis and classical psychometric tests in order to construct the health state classification. It then generates the health states using a Rasch-based approach developed by Mavranezouli and colleagues that avoids this problem (and presented in a companion paper) [9]. This paper takes the approach a stage further by showing how valuation of these states by the Time Trade-Off (TTO) preference elicitation technique can be used to estimate values for all states defined by the new health state classification.

The aim is to demonstrate the new approach and at the same time to derive a new preference-based measure for flushing symptoms. It presents the three key stages for deriving a preference-based measure: firstly deriving a health state descriptive system, illustrated using the Flushing Symptoms Questionnaire (FSQ) [10]; secondly the valuation survey; and thirdly the modelling the results from valuation survey and Rasch data to develop a method to estimate values for all states.

Methods

Flushing and the Flushing Symptoms Questionnaire (FSQ)
Flushing, also known as blushing, is a redness caused by increased blood flow. Although it can be unpleasant, it is not harmful or dangerous. Symptoms of flushing include an uncomfortable feeling of warmth or heat, prickling (itching or tingling sensation) and reddening of the skin anywhere on the body, but most commonly to the face, neck, chest or back. People may experience one symptom or several symptoms of flushing. Flushing can occur naturally, i.e. when a person is embarrassed, but is also a common and well recognised side effect of niacin medications – used to treat niacin deficiency, improve cholesterol levels or lower fat levels [11,12].

The FSQ was constructed to further characterise the symptoms of flushing as a side effect of taking niacin medications [10]. The questionnaire has been well validated. FSQ asks a series of 11 questions about flushing symptoms, of these seven ask respondents to rate their flushing symptoms on a 0 to 10 scale and as such are amenable for use in a PBM. The remaining four questions ask responders to confirm whether they took their study medication and ask about the frequency and length of flushing episodes.

The seven questions where respondents are asked to rate their symptoms relate to overall flushing symptoms, overall bother of flushing symptoms, redness of skin, warmth, tingling, itching and bother during the night. The authors of the questionnaire further suggest that responses on the ten-point scale can be categorised as: not at all bothersome (score 0), slightly bothersome (score 1 to 3), bothersome (score 4 to 6), very bothersome (score 7 to 9) and extremely bothersome (score 10) for the two bother related items and; did not have (score 0), mild (score 1 to 3), moderate (score 4 to 6), severe (score 7 to 9) and extreme (score 10) for the remaining five items.

FSQ is not currently a PBM, therefore the aim of this study was to derive a PBM based on the seven symptoms related questions. It is recognised that the FSQ is not multidimensional and the items of the FSQ correlate with each other, therefore traditional methods for selecting health states for valuation were not applicable.

Confirming the Unidimensionality of the HRQL Measure
In previous studies we have successfully used Rasch analysis [12] alongside traditional psychometric analysis to aid in the construction of health state classifications [6, 7]. These techniques will firstly be used here to validate whether the seven FSQ items unidimensionally measure flushing symptoms. The process of validating a HRQL measure using Rasch analysis has previously been described in detail [7] and interested readers are referred to this article for further details. A brief description of the process is summarised below.

Prior to Rasch modelling, Spearman's correlation and principal component factor analysis were used to establish the level of unidimensionality between the seven FSQ items [14]. In the principal component analysis Eigen values greater than 1 and scree plots were used to establish dimensionality and varimax component matrices were examined to establish the loading of items onto the dimensions identified by the Eigen values.

It is well recognised that Rasch analysis should be used alongside classical psychometric methods in the construction of health related quality of life (HRQL) instruments [15], therefore the flushing data were also fitted to Rasch partial credit models in the process of establishing the unidimensionality of the FSQ. This was a two step process that involved examining item level ordering and overall Rasch model goodness of fit statistics.

The first stage in validating a HRQL measure using Rasch modelling is to establish where responders are able to distinguish between the item level categories. Item threshold probability curves (a plot of the probability of being in each item level across the latent [HRQL] scale, depending on symptom severity) were examined for each item to identify items where responders to the FSQ were unable to distinguish between item response levels. If an item is ordered, the thresholds between item levels are the points where each item level is equally likely to occur. Disordered items highlight the inability of responders to distinguish between item levels. It should be noted that the objective of this study was not to change the wording of the FSQ but to derive a PBM that could be derived from the original measure. Therefore, we did not use formal Rasch guidelines for item level collapsing [16]. Instead, adjacent item levels for unordered items were merged, where combinations of item levels that gave the best Rasch model item-trait goodness of fit (see below for definition) were selected.

Once item level ordering has been achieved for all items the next stage is to establish unidimensionality under the Rasch model, by examining the overall Rasch model goodness of fit statistics and individual item goodness of fit statistics (the item-trait interactions and person and item fit residuals).

The **item-trait** interaction measures whether data fit the Rasch model for discrete groups of responders, where responders who have similar HRQL, based on their responses to the FSQ, are grouped together. The results are summarised using a χ^2 statistic, where a p-value for the overall model greater than 0.01 indicates a good fitting (unidimensional) Rasch model [16]. Item and person **fit residuals** examine the amount of variability between the expected and observed responses. The residuals are standardised so that the mean item or person fit statistic should be approximately zero with a standard deviation approximately equal to one.

If the overall goodness-of-fit of the data to the Rasch model was poor, suggesting non-unidimensionality, then individual item goodness of fit statistics were examined (item χ^2 statistics and item residuals) and any item with a large residual or poor χ^2 statistic was removed from the Rasch model. The process was repeated until only well fitting items remained and the overall item-trait goodness of fit of the model was non-significant (i.e. unidimensionality was achieved).

Item level ordering was re-examined at each stage of the Rasch analysis.

Using Rasch Analysis to Select Health States for Valuation
Rasch item threshold maps were used to derive health states amenable for the valuation stage of the development of a PBM, using the process described in the companion paper [9]. To summarise, the item threshold map, presented on the logit scale, can be used to create a set of health states across the items selected for inclusion in the PBM. For example (Figure 1), at logit -5 respondents are most likely to be in state 11111, indicating no symptoms of redness of skin, warmth, tingling and itching and not at all or slight difficulty sleeping. Whereas at logit 0 they are most likely to be in state 33331, indicating moderate symptoms of redness of skin, warmth, tingling and itching and not at all or slight difficulty sleeping. By moving from left to right across the item threshold map we can construct a set of plausible health states that are amenable for the valuation stage of creating the PBM. These states are logical and are based on the natural occurrence of states for people suffering from the underlying disease/condition e.g. flushing.

Alternative Methods of Selecting Health States
In order to illustrate the impracticalities of using traditional methods of selecting health states for the valuation stage of creating a PBM the health states selected using Rasch analysis were compared with those derived using an orthogonal block design as used on the SF-6D [4] and OAB-5D [18]. The number of states generated by this method was set so that the number was the same as that identified from the Rasch analysis.

The coverage of the selected health states created by Rasch modelling and orthogonal block design was compared using chi-squared statistics.

Valuation Survey
A valuation survey using face-to-face interviews was undertaken in order to determine people's preferences for the health states chosen using Rasch analysis. The main valuation technique used to value health states was TTO. Interviews were conducted by trained and experienced interviewers from the Centre for Research and Evaluation (CRE) at Sheffield Hallam University. Respondents were selected using sampling from streets in both urban and rural areas with a mix of socio-economic characteristics in the North of England. Households in these areas received letters informing them that interviewers would be in their area and interviewers then visited houses. All willing participants were then interviewed.

Interviews were undertaken in the respondent's own home. Respondents were not offered any financial reward for their participation. Respondents were asked firstly to complete the classification for their own health state to familiarise the respondent with the idea of describing states and the items and levels in the descriptive system. Respondents then undertook warm-up ranking and TTO tasks and eight TTO valuations of health states. The Measurement and Valuation of Health (MVH) group version of TTO was used to allow comparison with the EQ-5D tariff [2]. The TTO asks people to imagine they will be in a given health state for 10 years, and then asks them to consider a number of shorter periods in full health (x). At the point where respondents are unable to choose, the value of the state is given as x/10. Respondents were also asked a number of background questions covering health, demographic and socio-economic characteristics and how difficult they found the valuation tasks.

Two sets of cards for the valuation tasks were chosen based on health states generated from the results of the Rasch analysis. There were three versions of the interview; one for each set of cards and a third version using the same set of cards as the second version yet the generic full health card used in the TTO was the condition-specific description of full health (no symptoms of redness of skin, warmth, tingling and itching and not at all or slight difficulty sleeping) rather than the term 'full health'.

Using Rasch Analysis to create Health State Values for all States in the PBM

Mean TTO valuation scores were obtained from the valuation survey for each health state derived from the item threshold map. These valuation scores were mapped onto the Rasch model logit scale, using the logit values obtained from the Rasch model that generated the original health states. The relationship between valuation scores and Rasch logit values was examined and a series of regression models were fitted which included a simple linear relationship, quadratic, cubic, square root and cube root relationship. Model fit was compared using R^2 and root mean squared error (MSE) and the model that gave the best fit was selected as the most suitable model for predicting mean population-based valuation scores from Rasch logit scores.

The Flushing Symptom Questionnaire Dataset

Rasch models were applied to a dataset consisting of 1270 responders to the FSQ seven days after taking niacin medications. Our experience of valuation surveys has shown that respondents will struggle to distinguish between a large number of item levels. Therefore, rather than using the original ten point FSQ item scale, the categorised version of the scale was used in the construction of the PBM.

RUMM2020 [19] was used for all Rasch analysis, SPSS version 14 was used to generate health states for an orthogonal design [20] and STATA version 10.0 for all other statistical analysis [21].

Results

Table 1 presents Spearman's correlation matrix for the seven items amenable for inclusion in a preference-based flushing symptoms measure. The table shows high levels of correlation between the seven items. Principal component analysis confirmed that all seven items belonged to the same dimension (Figure 2), where the eigen vale for principal component 1 was almost 7 and those for principal components 2 to 7 were close to 0. Further, the factor loadings for principal component 1 were all extremely high (> 0.98) supporting the items belonging to a uni-dimensional instrument.

A Rasch model was fitted to the 1270 responders to the seven FSQ items and Rasch analysis was first used to check whether respondents could distinguish between different item levels for each item (item level ordering). Item threshold curves were examined for each item; all items except sleep difficulty were ordered. For the sleep item, item level "slightly bothersome" could either be merged with "not at all bothersome" or "bothersome." Two Rasch analysis were run to explore which combination gave the better fitting Rasch model, the first merged levels for "not at all bothersome" and "slightly bothersome" together (χ^2_{49} = 156.84, p < 0.001) and the second merged levels for "slightly bothersome" and "bothersome" together (χ^2_{56} = 228.54, p < 0.001). The Rasch model merging "not at all" and "slightly bothersome" gave a lower (better) model goodness of fit statistic and was thus selected as the most appropriate merging of item levels for sleep bother item.

The next stage in deriving a unidimensional Rasch model is to establish whether all items fit the Rasch model by examining the overall model and item goodness of fit statistics. The Rasch model that included all 7 FSQ items had a significant chi-squared statistic and an overall mean item residual with a high standard deviation (mean = -1.07, standard deviation = 3.79) suggests that one or more items could be removed from the model, as they do not fit it well. Table 2 presents the item fit statistics and residuals for the seven FSQ items. Examination of item fit statistics suggested that four items did not fit the Rasch model: overall symptoms, overall bother, tingling and itching. Further, overall symptoms and overall bother had high negative fit residuals, of the two items overall symptoms had the highest residual and chi-squared statistic and was removed from the Rasch model.

Removal of the item asking about overall symptoms improved the overall goodness of fit of the remaining items to the Rasch model, although the model fit was still significant (χ^2_{42} = 102.65, p < 0.001, mean = -0.70, standard deviation = 2.06) indicating further items could be removed from the model. The individual item goodness of fit and residuals were again examined and this time only the overall bother item failed to fit the Rasch model (Table 2). As before, this item was removed, the Rasch model was refitted and overall model goodness of fit was achieved (χ^2_{30} = 44.64, p = 0.042, mean = -0.70, standard deviation = 1.14). Therefore, the final health state classification to be developed into a PBM contained five items asking about redness of skin, warmth, tingling, itching and sleep difficulty (Figure 3).

The item threshold map was examined for the five remaining FSQ items, and this identified a total of 16 possible health states ranging from condition specific full health (11111) to worst possible health with flushing symptoms ("PITS" state 55554) (Table 3). These states cover 76% of responses to the FSQ for the 1270 responders and 32% of the 447 who responded excluding those responding in condition specific full health (11111).

Table 3 also contains 16 health states generated by orthogonal design; the orthogonally selected states cover 67% of responses to the FSQ, this is a significantly lower proportion of coverage than those selected using the Rasch method (χ_1^2 = 28.66, $p < 0.001$). Only three of these states, condition specific full health (11111), 12221, and PITS state (55554) overlap between both selection processes, these three states cover 67% of the 1270 responses to the FSQ and after excluding full health cover only 5% of responses. The 13 states selected from the orthogonal design that are not common with the Rasch selected states do not cover any further responders to FSQ, whereas the 13 Rasch states that were not common to the orthogonal selection cover a further 10% of responses.

Valuation Survey

There were 219 successfully conducted interviews to value the 16 states, a response rate of 44.7% for respondents answering their door at time of interview. The study achieved a completion rate of 99.9% for all states included in the TTO valuations (only 1 missing value). Two respondents were excluded for valuing all states as identical and less than 1. One respondent was excluded for valuing the PITS state higher than every other state and one respondent was excluded for valuing all states as worse than dead. Characteristics of all respondents included in the analysis are presented in Table 4 and compared to the general population in South Yorkshire and England. The sample has a higher proportion of retired individuals, home owners and females and a lower proportion of unemployed individuals. The interviewers reported that it was doubtful whether the respondent understood the TTO tasks for less than 2% of all respondents.

Mean health state values

Table 5 reports descriptive statistics for the health state values. The mode (most frequently reported value) for 15 of the 16 health states valued is 1, which may be considered reasonable for a condition specific measure with a descriptive system that does not contain dimensions such as pain or mobility. Ordinary least squares (OLS) regression models were fitted to account for any differences explained by personal characteristics of the people valuing the health states. Table 6 reports summary statistics for the predicted health state values after adjusting for the levels of the health state being valued, age, gender, marital status, employment status and home ownership of the people valuing the health states using OLS.

Modelling health state utility values from Rasch logit values

Logit values can be obtained for each respondent included in the Rasch model and these values relate to a specific set of item responses in the flushing questionnaire. For example, respondent A, with Rasch logit -5.275 responded that they had no redness of skin, warmth, tingling, itching as a result of flushing and difficulty sleeping as a result of flushing that "was not at all or slightly bothersome" (questionnaire wording). Logit values were rescaled and anchored at 0.98 for condition specific full health (state 11111) and 0.42 for worst health (PITS state 55554) which were the mean values given by responders in the valuation survey.

Figure 4 presents the relationship between the mean health state utility values and the rescaled logit values derived from responders based on the Rasch analysis. It was noticed that there was a large gap in the mean utility values between states including no symptoms, mild or moderate symptoms (levels 1, 2 or 3) and states including severe or extreme symptoms (levels 4 or 5) (See Table 5 & 6 valuation values). Therefore a series of models were hypothesised which included a dummy variable to allow for the "jump" in population values between mild and moderate states. Models ranged from a simple linear relationship to quadratic, cubic, square root and cube root. The justification for exploring these relationships was that the examination of points in Figure 4 suggested that the valuation scores followed a slight curve. The fitted models are listed below:

Model 1: Linear relationship – $y = \text{Constant} + b_1x + b_2\text{Dummy}$

Model 2: Quadratic relationship – $y = \text{Constant} + b_1x^2 + b_2\text{Dummy}$

Model 3: Quadratic relationship – $y = \text{Constant} + b_1x^2 + b_2x + b_3\text{Dummy}$

Model 4: Cubic relationship – $y = \text{Constant} + b_1x^3 + b_2\text{Dummy}$

Model 5: Cubic relationship – $y = \text{Constant} + b_1x^3 + b_2x^2 + b_3\text{Dummy}$

Model 6: Cubic relationship – $y = \text{Constant} + b_1x^3 + b_2x + b_3\text{Dummy}$

Model 7: Cubic relationship – $y = \text{Constant} + b_1x^3 + b_2x^2 + b_3x + b_4\text{Dummy}$

Model 8: Square root relationship – $y = \text{Constant} + b_1x^{1/2} + b_2\text{Dummy}$

Model 9: Cubic root relationship – $y = \text{Constant} + b_1x^{1/3} + b_2\text{Dummy}$

where y = mean valuation score, x = Rasch (rescaled) logit score, b_i = regression coefficient and Dummy = dummy variable for moderate states.

Table 7 shows that all models had very good goodness of fit with R^2 statistics all greater than 0.9. Models 3, 6, 7, 8 and 9 had a non-significant regression term and were thus excluded from consideration. Of the remaining models, all had similar goodness of fit statistics, given this we are recommending the simplest model (Model 1) as the model to use to derive TTO values for non valued states based on Rasch model scores. This model can be used to derive health state values for all potential health states, including those not assessed in the valuation survey. As an example, using model 1 with a dummy coefficient (Table 7) a person in state 11112, indicating no symptoms of redness of skin, warmth, tingling or itching with bothersome sleeping, with rescale Rasch logit value of 0.98 would have a valuation value of 1.00. Further, a person in state 34221, indicating moderate redness of skin, severe warmth, mild tingling and itching and not at all or slightly bothered during sleep as a result of flushing, with Rasch (rescaled) logit value of 0.678 would have a TTO valuation estimate of 0.627.

Discussion

This paper has applied a new approach to developing a preference-based measure from an existing non-preference-based measure of health. Having derived a health state classification from the Flushing Symptoms Questionnaire, it uses a new Rasch based method for selecting states for valuation. For the first time, it estimates the relationship between the Rasch latent variable and TTO–based health state utility values in order to estimate values for all states described by the health state classification. This paper demonstrates how this new approach can be used to generate a preference-based index from existing measures where the items are highly correlated.

A potential disadvantage with the Rasch-based vignette approach, like clustering, is that it generates far fewer states than those described by the health state classification. We have proposed a method to solve this problem by estimating a relationship between the points on the preference scale and the latent variable produced by the Rasch FSQ model using a simple regression function. This enables the estimation of preference values for other points on the latent variable and hence other states generated using the health state classification.

There is a more general interest in understanding the relationship between utilities and methods such as Rasch and item response theory (IRT). Utility theory and Rasch model concepts come from very different traditions and perhaps in the past have been seen as competing techniques. However, this paper shows how the techniques can be used in tandem. They also show the high correlation between the Rasch model logit scale and the utility scale, though the relationship found in this example indicates that one can not be used entirely as a substitute for the other.

This use of the FSQ to develop a preference-based measure could be criticised for being unusually narrow and so most prone to the problem of high correlation between items. However, as found in the companion paper [9], this approach can also be used in the context of a broader measure of psychological health. Additionally, it may well have applications for other condition specific measures, focusing on difficulties associated with problems of vision or hearing. Further research is needed to explore the generalisability of this new approach. It might also be argued that perhaps unidimensional health states, such as the FSQ, should be based on single items and so avoid the problem of working with multi-item instruments. The arguments for single item scales are well known in terms of improved reliability [22]. Another argument is that the selection of multiple items from a unidimensional instrument provides a much richer description for members of the public trying to imagine the health state for valuation than a single item description would do.

There are more general concerns about the use of condition specific preference-based measure that are beyond the scope of this paper. Concerns have been expressed in the literature about the appropriateness of condition specific measures for use in making cross programme comparison including the impact of side effects, co-morbidities, the appropriate anchors and achieving comparability across programmes. These issues have been addressed in more detail elsewhere. [5]

Condition specific measures may not always have independent items, and existing techniques of developing health state values from condition specific measures are inappropriate when items are not independent. We have applied an alternative technique using Rasch analysis that is appropriate under these circumstances, which avoids implausible health states generated by statistical designs (e.g. an orthogonal array) and then shown how the valuation of these states can be mapped onto the Rasch scale in order to value all states defined by the instrument. This should significantly enhance work in this field.

References

1. Drummond MF, Sculpher MJ, O'Brien BJ, Stoddart GL, Torrance GW (2005). Methods for the economic evaluation of health care programmes (3rd Edition).
Oxford University Press, Oxford

2. Dolan P (1997).
Modelling valuations for EuroQol Health States.
Medical Care 35(11):1095-108.

3. Feeny D, Furlong W, Torrance G, Goldsmith C, Zenglong Z, DePauw S, Denton M, Boyle M (2002).
Multiattribute and single attribute utility functions for the Health Utilities Index Mark 3 system.
Medical Care 40:113-28.

4. Brazier JE, Deverill M, Harper R, Booth A (1999).
A review of the use of Health Status measures in economic evaluation.
Health Technol Assess 3(9).

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

6. Young T, Yang Y, Brazier J, Tsuchiya A (2007).
The use of Rasch analysis as a tool in the construction of a preference-based measure: the case of AQLQ.
HEDS Discussion Paper Series No. 07/01
http://www.shef.ac.uk/scharr/sections/heds/discussion.html

7. Young T, Yang Y, Brazier J, Tsuchiya A, Coyne K (2009).
The first stage of developing preference-based measures: constructing a health state classification using Rasch Analysis.
Quality of Life Research 18(2):253-65.

8. Brazier JE, Harper R, Thomas K, Jones N, Underwood T (1998).
Deriving a preference-based single index measure from the SF-36.
J Clinical Epidemiol 51(11):1115-29.

9. Mavranezouli I, Brazier JE, Young TA, Barkham M.
Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-5D Utility from CORE-OM in order to elicit preferences for common mental health problems.
(In submission, 2009)

10. Norquist J, Watson DJ, Yu Q Winfen, Paolini JF, McQuarrie K, Santanello NC (2007).
Validation of a questionnaire to assess niacin-induced cutaneous flushing.
Current Medical Research and Opinion 23(7):1549-60.

11. AHFS Consumer Medication Information. American Society of Health-Systems Pharmacists.
http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=meds&log$=drug_bottom_one&part=a682518 Last visited on 07/10/2009

12. Bodor ET, Offermanns S (2008).
Nicotinic Acid: An old drug with a promising future.
British Journal of Pharmacology 153(S1): S68-S75.

13. Rasch G (1960).
Probabilistic models for some intelligence and attainment tests.
Chicago: University of Chicago Press: Reprinted 1980.

14. Chatfield C, Collins AJ (1980).
Introduction to Multivariate Analysis.
Chapman and Hall; University Press, Cambridge.

15. Tennant A, McKenna SP, Hagell P (2004).
Application of Rasch analysis in the development and application of quality of life instruments.
Value in Health 7(Supplement 1):S22-S26.

16. Linacre JM (1999).
Investigating rating scale category utility.
Journal of Outcome Measurement 3(2):103-22.

17. Kubinger KD (2005).
Psychological test calibration using the Rasch model – Some critical suggestions on traditional approaches.
International Journal of Testing 5(4): 377-94.

18. Yang Y, Brazier J, Tsuchiya A, Coyne K (2008).
Estimating a preference-based single index from the Overactive Bladder questionnaire (OAB-q).
Value in Health 12(1):159-66.

19. Rasch Unidimensional Measurement Models (RUMM) 2020 ©. RUMM Laboratory Pty Ltd 1997-2004

20. SPSS Inc. Release 14.0.1 Chicago, Illinois; SPSS Inc. 2005

21. StataCorp. Statistical Software: Release 10.0. College Station, Texas: Stata Corporation. 2008

22. Youngblut JM, Casper GR (1993).
Single-item indicators in nursing research.
Research in Nursing and Health 16(6):459-65.

Table 1: Spearman's correlation matrix (N = 1270)

	Overall Symptoms	Overall Bother	Redness	Warmth	Tingling	Itching	Sleep Difficulty
Overall Symptoms	1.00	0.96	0.83	0.90	0.81	0.84	0.69
Overall bother		1.00	0.82	0.88	0.80	0.84	0.70
Redness of skin			1.00	0.84	0.77	0.78	0.71
Warmth				1.00	0.78	0.77	0.71
Tingling					1.00	0.80	0.69
Itching						1.00	0.70
Sleep Difficulty							1.00

Table 2: FSQ Rasch Model Item Residual and Goodness of Fit Statistics

	Rasch Model: All 7 FSQ items included			Rasch Model: Excluding overall symptoms item			Rasch Model: Excluding overall symptoms item and overall bother item		
	Residual	χ^2_7	P-value	Residual	χ^2_7	P-value	Residual	χ^2_7	P-value
Overall Symptoms	-6.50	37.74	< 0.01						
Overall Bother	-5.63	28.12	< 0.01	-4.06	38.43	< 0.01			
Redness of skin	-0.39	16.48	0.02	-1.09	23.05	< 0.01	-2.42	17.44	0.01
Warmth	-0.10	2.32	0.94	-0.52	5.40	0.61	-0.46	3.23	0.78
Tingling	3.93	28.12	< 0.01	2.21	13.46	0.06	0.60	6.17	0.40
Itching	2.00	38.54	< 0.01	0.36	14.10	0.05	-0.01	8.98	0.17
Sleep Difficulty	-0.78	5.54	0.60	-1.11	8.21	0.31	-1.27	8.83	0.18

Table 3: All possible logical states as shown in the Rasch item threshold map and 16 states (including best health and pits) selected assuming orthogonal (independent) design

States identified by Rasch analysis (% of sample in selected state)		States selected by orthogonal design (% of sample in selected state)	
1	11111 (65%)	1	11111 (65%)
2	12111 (3%)	2	12221 (2%)
3	12121 (1%)	3	14443 (0.0%)
4	12221 (2%)	4	21232 (0.0%)
5	22221 (4%)	5	24514 (0.0%)
6	22231 (0.5%)	6	25123 (0.0%)
7	32231 (0.1%)	7	31353 (0.0%)
8	33231 (0.0%)	8	32412 (0.0%)
9	33331 (0.7%)	9	33521 (0.0%)
10	33342 (0.0%)	10	34131 (0.0%)
11	43343 (0.1%)	11	41424 (0.0%)
12	43443 (0.1%)	12	44251 (0.0%)
13	44443 (0.2%)	13	52154 (0.0%)
14	44453 (0.0%)	14	53213 (0.0%)
15	54453 (0.0%)	15	55432 (0.0%)
16	55554 (0.1%)	16	55554 (0.1%)

Table 4: Characteristics of respondents to the valuation survey compared with population characteristics for South Yorkshire and England

	Included respondents (n=215)	South Yorkshire[1]	England
Mean age (standard deviation)	44.92 (18.37)	-	-
Female	53.5%	51.2%	51.3%
Married/Partner	63.7%	-	-
Employed or self-employed	51.6%	56.1%	60.9%
Unemployed	0.9%	4.1%	3.4%
Long-term sick	4.2%	7.7%	5.3%
Full-time student	9.3%	7.5%	7.3%
Retired	22.3%	14.4%	13.5%
Own home outright or with a mortgage	70.1%	64.0%	68.7%
Renting property	29.9%	36.0%	31.3%
Secondary school is highest level of education	33.0%	-	-
EQ-5D score (standard deviation)	0.85 (0.22)	-	-
Found rank valuation task difficult (judged by respondent)	14.5%	-	-
Found TTO valuation task difficult (judged by respondent)	18.8%	-	-
Doubtful whether the respondent understood the rank task (judged by interviewer)	2.3%	-	-
Doubtful whether the respondent understood the TTO tasks (judged by interviewer)	1.4%	-	-
TTO completion rate	99.9%		

[1] Statistics for South Yorkshire Health Authority and for England in the Census 2001. Questions used in this study and the census are not identical. The census includes persons aged 16 and above whereas this study only surveys persons aged 18 and above.

Table 5: Mean TTO values by health state

	Count	Mean	Minimum	Maximum	25th Percentile	Median	75th Percentile	Standard Deviation	Mode
11111	75	0.98	0.38	1.00	1.00	1.00	1.00	0.10	1.00
12111	140	0.94	0.03	1.00	0.98	1.00	1.00	0.16	1.00
12121	75	0.95	0.28	1.00	0.93	1.00	1.00	0.12	1.00
12221	140	0.89	0.03	1.00	0.83	1.00	1.00	0.19	1.00
22221	75	0.89	0.03	1.00	0.88	1.00	1.00	0.21	1.00
22231	140	0.87	0.28	1.00	0.80	0.95	1.00	0.19	1.00
32231	75	0.87	0.03	1.00	0.83	0.99	1.00	0.22	1.00
33231	140	0.83	-0.48	1.00	0.73	0.93	1.00	0.26	1.00
33331	75	0.84	0.03	1.00	0.78	0.99	1.00	0.24	1.00
33342	140	0.68	-0.73	1.00	0.53	0.75	0.93	0.31	1.00
43343	140	0.57	-1.00	1.00	0.33	0.63	0.93	0.40	1.00
43443	75	0.60	-0.43	1.00	0.38	0.68	0.93	0.34	1.00
44443	140	0.47	-1.00	1.00	0.20	0.50	0.83	0.41	1.00
44453	75	0.54	-0.38	1.00	0.23	0.57	0.88	0.36	1.00
54453	75	0.52	-0.33	1.00	0.28	0.50	0.83	0.34	.00
55554	140	0.41	-1.00	1.00	0.13	0.43	0.78	0.45	1.00

Table 6: Modelled TTO values by health state adjusting for health state levels and personal characteristics of responders

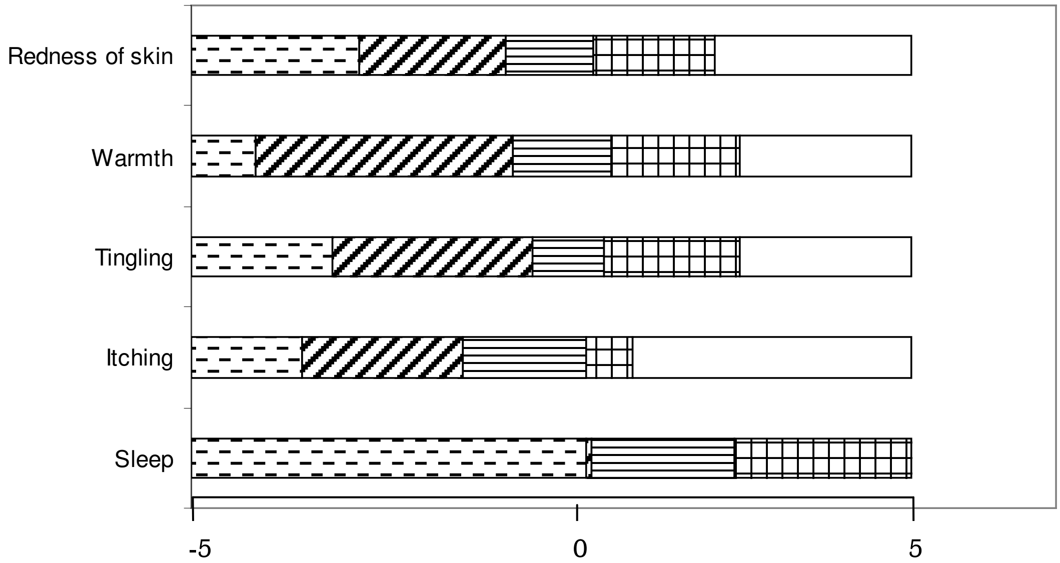
	Count	Mean	Minimum	Maximum	25th Percentile	Median	75th Percentile	Standard Deviation	Mode
11111	75	0.98	0.82	1.09	0.95	1.00	1.03	0.70	0.91
12111	72	0.95	0.76	1.06	0.92	0.96	1.00	0.06	0.93
12121	75	0.95	0.79	1.06	0.92	0.97	1.00	0.07	0.88
12221	72	0.90	0.72	1.01	0.87	0.91	0.95	0.06	0.89
22221	75	0.89	0.73	1.00	0.86	0.91	0.94	0.07	0.82
22231	72	0.87	0.69	0.98	0.84	0.89	0.92	0.06	0.86
32231	75	0.86	0.70	0.98	0.83	0.88	0.91	0.07	0.79
33231	72	0.83	0.65	0.94	0.80	0.85	0.88	0.06	0.82
33331	75	0.84	0.68	0.95	0.81	0.86	0.89	0.07	0.77
33342	72	0.69	0.50	0.79	0.65	0.70	0.73	0.06	0.67
43343	72	0.58	0.39	0.69	0.55	0.59	0.63	0.06	0.56
43443	75	0.60	0.44	0.71	0.57	0.61	0.64	0.07	0.53
44443	72	0.47	0.29	0.58	0.44	0.49	0.52	0.06	0.46
44453	75	0.54	0.38	0.65	0.51	0.55	0.58	0.07	0.47
54453	75	0.51	0.35	0.63	0.49	0.53	0.56	0.07	0.45
55554	72	0.42	0.24	0.53	0.39	0.43	0.47	0.06	0.41

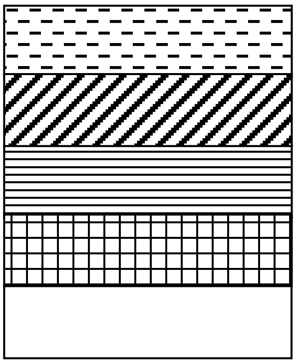
Table 7: Model 1 to 9 regression results and goodness of fit statistics for predicting mean valuation states from Rasch (rescaled) logit values after adding dummy variables to allow for severe states

		Constant	X^3	X^2	X	Dummy	R^2	Root MSE
Model 1	y = Constant + b_1x + b_2Dummy	0.42 (0.093)			0.60 (0.117)	-0.20 (0.036)	0.958	0.042
Model 2	y = Constant + b_1x^2 + b_2Dummy	0.66 (0.058)		0.37 (0.088)		-0.23 (0.038)	0.946	0.048
Model 3	y = Constant + b_1x^2 + b_2x + b_3Dummy	0.01 (0.232)^{ns}		-0.77 (0.403)^{ns}	1.73 (0.606)	-0.18 (0.035)	0.968	0.039
Model 4	y = Constant + b_1x^3 + b_2Dummy	0.75 (0.046)	0.28 (0.081)			-0.25 (0.039)	0.935	0.053
Model 5	y = Constant + b_1x^3 + b_2x^2 + b_3Dummy	0.37 (0.114)	-1.26 (0.449)	1.86 (0.535)		-0.17 (0.037)	0.968	0.039
Model 6	y = Constant + b_1x^3 + b_2x + b_3Dummy	0.12 (0.183)^{ns}	-0.37 (0.195)^{ns}		1.23 (0.350)	-0.18 (0.036)	0.968	0.039
Model 7	y = Constant + b_1x^3 + b_2x^2 + b_3x + b_4Dummy	0.03 (0.882)^{ns}	-0.06 (3.12)^{ns}	-0.65 (6.44)^{ns}	1.65 (4.24)^{ns}	-0.18 (0.044)	0.968	0.040
		Constant	$X^{1/2}$			**Dummy**	R^2	**Root MSE**
Model 8	y = Constant + $b_1x^{1/2}$ + b_2Dummy	-0.02 (0.164)^{ns}	1.04 (0.184)			-0.19 (0.035)	0.963	0.040
		Constant	$X^{1/3}$			**Dummy**	R^2	**Root MSE**
Model 9	y = Constant + $b_1x^{1/3}$ + b_2Dummy	-0.46 (0.235)^{ns}	1.47 (0.255)			-0.19 (0.034)	0.965	0.039

ns = Non-significant coefficient

Figure 1: Item Threshold Map





	Redness of skin, warmth, tingling, itching	Sleep Difficulty
Level 1	Did not have	Not at all or slightly bothersome
Level 2	Mild	Bothersome
Level 3	Moderate	Very bothersome
Level 4	Severe	Extremely bothersome
Level 5	Extreme	

Figure 2: Results of Principal Component Analysis (Scree Plot)

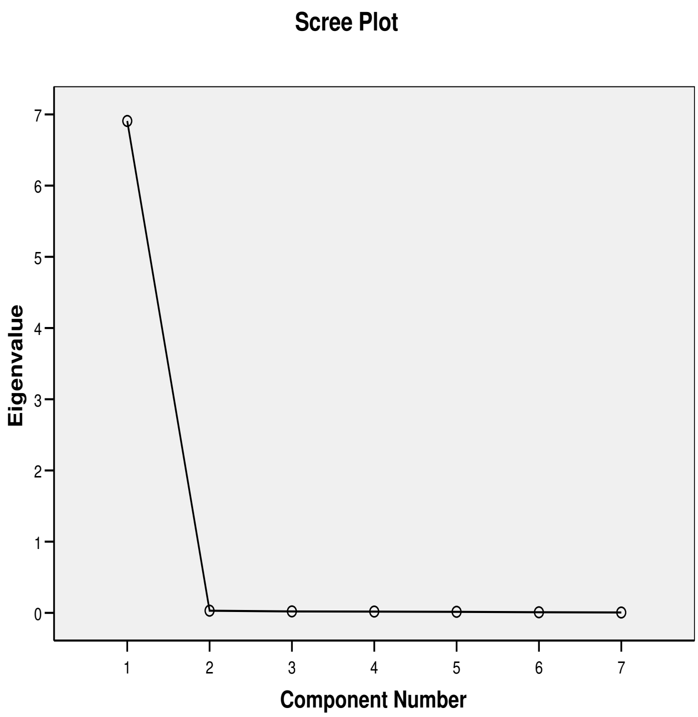


Figure 3: Flushing Symptoms Health State Classification

Redness of skin
1. No redness of skin as a result of flushing
2. Mild redness of skin as a result of flushing
3. Moderate redness of skin as a result of flushing
4. Severe redness of skin as a result of flushing
5. Extreme redness of skin as a result of flushing

Warmth
1. No warmth as a result of flushing
2. Mild warmth as a result of flushing
3. Moderate warmth as a result of flushing
4. Severe warmth as a result of flushing
5. Extreme warmth as a result of flushing

Tingling
1. No tingling as a result of flushing
2. Mild tingling as a result of flushing
3. Moderate tingling as a result of flushing
4. Severe tingling as a result of flushing
5. Extreme tingling as a result of flushing

Itching
1. No itching as a result of flushing
2. Mild itching as a result of flushing
3. Moderate itching as a result of flushing
4. Severe itching as a result of flushing
5. Extreme itching as a result of flushing

Sleeping
1. No difficulty sleeping as a result of flushing, or difficulty sleeping as a result of flushing is not at all bothersome or slightly bothersome
2. Difficulty sleeping as a result of flushing is bothersome
3. Difficulty sleeping as a result of flushing is very bothersome
4. Difficulty sleeping as a result of flushing is extremely bothersome

Figure 4 Scatter plot of relationship between Rasch rescaled logit values and mean valuation scores with regression line indicating the linear relationship (including a dummy variable for the gap between 0.8 and 0.6 on the valuation scale (N = 16)