



ELSEVIER

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Mapping EQ-5D Utility Scores from the Incontinence Quality of Life Questionnaire among Patients with Neurogenic and Idiopathic Overactive Bladder

Stephen Kay, MSc¹, Keith Tolley, BA, MPhil², Danielle Colayco, PharmD, MS³, Kristin Khalaf, PharmD, MS³, Peter Anderson, BSc^{1,*}, Denise Globe, PhD³

¹Adelphi Real World, Macclesfield, UK; ²Tolley Health Economics Ltd., Buxton, UK; ³Allergan, Inc., Irvine, CA, USA

ABSTRACT

Objectives: To provide a mapping algorithm for estimating EuroQol five-dimensional (EQ-5D) questionnaire index scores from the Incontinence-specific Quality of Life questionnaire (I-QOL) based on nationally representative samples of patients with idiopathic or neurogenic overactive bladder (OAB) using EQ-5D questionnaire preference valuations based on both the UK and US general populations. **Methods:** Analyses were conducted for 2505 patients from the Adelphi Overactive Bladder Disease Specific Programme, a cross-sectional study of patients with idiopathic or neurogenic OAB, undertaken in the United States and Europe in 2010. A range of statistical modeling techniques was used. Tenfold cross-validation techniques were used to calculate mean absolute error (MAE) and root mean squared error (RMSE) goodness-of-fit statistics. Various predictor lists, together with a method combining stepwise selection with multivariable fractional polynomial techniques to allow nonlinear relationships to feature, were pursued. **Results:** Choice of predictors was consistent for both the UK and US EQ-5D

questionnaire tariffs. For idiopathic, the best model included the I-QOL total score and age (both modeled nonlinearly.) For neurogenic, the best model was the I-QOL social embarrassment domain score modeled linearly only. Best-fit results were better in the idiopathic ($n = 2351$; MAE = 0.10; RMSE = 0.14) than in the neurogenic sample ($n = 254$; MAE = 0.17; RMSE = 0.22). **Conclusions:** This research provides algorithms for mapping EQ-5D questionnaire index scores from the I-QOL, allowing calculation of appropriate preference-based health-related quality-of-life scores for use in cost-effectiveness analyses when only I-QOL data are available. The strongest results were for idiopathic patients, but those for neurogenic are consistent with those of other published mapping studies.

Keywords: cross-walk, EQ-5D, I-QOL, mapping, real world, utility values.

Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

Utilities reflect the strength of an individual's preference for specific health-related outcomes and are commonly used to generate health state values for calculating quality-adjusted life-years. A single summary score is generated by applying societal preference weights to a health state classifier, completed by the patient, that ranges from 0 to 1 on an interval scale, where 9 represents a state equivalent to death and 1 reflects perfect health [1,2]. Such utility scores may be directly elicited by using preference-based techniques (e.g., standard gamble, time-trade off, or rating scales) completed by a representative sample of a regional general population. Alternatively, they can be measured indirectly through the use of algorithms applied to either generic or disease-specific questionnaires initially generated from direct elicitation by using preference-based techniques. When direct elicitation is not feasible, algorithms to estimate indirect utilities from health-related quality-of-life (HRQOL) questionnaires may also be developed by mapping an HRQOL questionnaire onto the utility algorithm of a generic instrument [3]. The ability to

translate from health status measures to preference utility scores is appealing in terms of minimizing respondent burden and analyzing data for which only health status measures are available.

The EuroQol five-dimensional (EQ-5D) questionnaire, one of the most commonly used generic questionnaires for deriving utility scores, is composed of five dimensions (mobility, self-care, usual activities, pain or discomfort, and anxiety or depression), each with three levels (no problems, some problems, or extreme problems/unable to). A total of 243 health states can be created, with 0 representing death and 1 representing perfect health. It has been widely used owing to its reported validity and reliability [4–6]. Country-specific value sets (community preference tariffs) exist for several countries, including the United Kingdom [7] and the United States [8]. A five-level version of the EQ-5D questionnaire is available; however, the associated tariff is interim and to date this version is not being routinely used in clinical studies [9].

Overactive bladder (OAB) syndrome is a symptom-based diagnosis and is defined as urgency, with or without urgency incontinence, usually with frequency and nocturia [10]. These

* Address correspondence to: Peter Anderson, Adelphi Real World Ltd., Adelphi Mill, Bollington, Cheshire SK10 5JB, UK.

E-mail: peter.anderson@adelphigroup.com.

1098-3015/\$36.00 – see front matter Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Published by Elsevier Inc.

<http://dx.doi.org/10.1016/j.jval.2012.12.005>

symptoms are a consequence of the bladder's inability to effectively store urine because of an underlying dysfunction in the coordinated mechanisms that provide controlled storage and voiding of urine. Patients with underlying neurologic conditions (e.g., multiple sclerosis, spinal cord injury) often experience neurological deficits or lesions within regions of the central nervous system that govern bladder function. As a result, symptoms consistent with OAB are commonly experienced in these patient subgroups and these patients are referred to as having neurogenic OAB. In the absence of a known neurological insult or pathophysiological cause, the reason for the dysfunction is unknown in the majority of cases, resulting in the condition referred to as idiopathic OAB.

OAB is a prevalent disorder that is reported to affect between 12% and 17% of the general population in North America and Europe [11–14]. The negative impact of OAB and urinary incontinence has been well established. The symptoms of OAB, including urinary incontinence, have psychosocial, physical, and sexual effects that impact patients' HRQOL [15–18]. OAB symptoms and urinary incontinence can have an impact on emotions (e.g., distress, embarrassment, and self-esteem), physical activities (e.g., difficulty with daily activities due to leakage), and social activities (e.g., restriction of activities due to anxiety about not being able to reach a toilet). In addition, incontinence and frequency of urination at night have a deleterious impact on a patient's ability to sleep, leading to daytime sleepiness and decreased energy level. Thus, OAB affects the individual's ability to function physically and mentally and is associated with symptoms of depression [12,18,19].

The Incontinence Quality of Life Questionnaire (I-QOL) is a disease-specific, patient-reported instrument initially designed to measure the impact of urinary incontinence on patients' lives. It has demonstrated sound measurement properties in prior studies among patients with urinary incontinence, including those who have OAB both with and without urinary incontinence [20–23]. The I-QOL consists of 22 items divided into three domains: avoidance and limiting behavior, psychosocial impact, and social embarrassment. Scores can be calculated for each domain, and a total summary score can also be calculated from all 22 items, with 100 representing the best possible score [24]. Since its development, the I-QOL has been and continues to be used in a number of clinical trials and other research studies, and its psychometric properties have been documented in a number of validation studies, showing it to be a valid and reliable measure in patients with stress incontinence, OAB, and urinary incontinence due to an underlying neurologic condition [25–29].

This article describes the modeling techniques used to map I-QOL onto the EQ-5D questionnaire for patients with OAB due to idiopathic or neurogenic etiologies. The study was restricted to patients diagnosed with OAB with or without urinary incontinence. Methods and results are presented separately for patients with idiopathic and neurogenic etiologies to account for the inherent variability between these two populations.

Methods

Data Source

Data were drawn from the Adelphi Overactive Bladder Disease Specific Programme (OAB DSP), a multinational, cross-sectional study of real-world clinical practice conducted in the United States and four European Union countries (France, Germany, Spain, and the United Kingdom) between October 2010 and January 2011. A total of 259 primary care physicians and 445 specialists (urologists, gynecologists, and urogynecologists) participated in the program. Physicians completed a patient record

form for the next 10 consecutive consulting patients who met the eligibility criteria, and the same patients were invited to fill out a patient self-completion form, which included the I-QOL and the EQ-5D questionnaire. All patients with symptoms of OAB/incontinence, as diagnosed by their physician, were eligible for inclusion in the survey except for those younger than 18 years, those with a lower urinary tract infection, and pregnant female patients. The real-world design of the study ensured that information available to the physician/patient only at the time of the consultation was collected. Therefore, no tests or investigations were required for a patient to be included in the study, nor were they conducted as part of the study itself. This methodology has been successfully administered for other disease states including respiratory, central nervous system, cardiovascular, oncology, and autoimmune conditions and has been outlined in detail previously [30].

Analysis Plan

To address the specific objectives outlined in this article, the following patients were excluded from the sample prior to analysis:

- Patients with stress urinary incontinence only (but included mixed urinary incontinence)
- Patients who had surgery for OAB (procedures to insert indwelling catheters or inject bulking agents were not excluded)
- Patients whose primary neurogenic condition was reported as a complete spinal cord injury, stroke, Parkinson's disease, or "Other"
- Patients with concomitant cancer or benign prostatic hyperplasia.

The first two criteria were applied to minimize variability within the general OAB/incontinence population. The third criterion attempted to restrict the population to patients with either idiopathic OAB or those with a specific subset of neurologic conditions. The final criterion excluded patients whose quality-of-life decrement may be attributed to conditions other than their urinary incontinence. The final sample consisted of patients with idiopathic OAB and patients with a primary neurologic condition of either multiple sclerosis or incomplete spinal cord injury who experienced symptoms of OAB that were thought to stem from neurogenic detrusor overactivity. Given the inherent differences between these two groups, the analysis was split by disease type—neurogenic and idiopathic OAB. Results of the idiopathic OAB group are presented first, followed by the results for patients with neurogenic OAB.

Statistical Methods

A number of statistical models have historically been used to map (cross-walk) various health status measures onto EQ-5D questionnaire community preference values. Traditional mapping methods include ordinary least squares (OLS), Tobit, censored least absolute deviation (CLAD), generalized linear models (GLMs), and two-part models (2PM). Given the variety of models available, characteristics of the EQ-5D questionnaire were taken into consideration to determine which models would provide viable options to test for this mapping exercise.

An important factor to consider is the interpretation of a perfect score on the EQ-5D questionnaire (i.e., a score of 1). There is a ceiling score of 1, which means that the EQ-5D questionnaire should be considered as censored in that while the modal EQ-5D questionnaire score is typically 1, no individual observation can exceed 1. Furthermore, the modal score of 1 (perfect health) can be a direct result of the EQ-5D questionnaire being an instrument

of general well-being, and as such may be insensitive to nuances of specific diseases and of insufficient granularity to capture disease-specific impacts [31]. In this instance, OLS estimates would be biased; however, both CLAD and Tobit models could be appropriate candidates. CLAD models are similar to Tobit models but are superior in the presence of heteroskedasticity or nonnormality [32]. GLMs, however, are very flexible models that can represent many different forms of distributions. Dependent on the shape of the EQ-5D questionnaire distribution, appropriate GLM link (e.g., log, identity) and distribution families (e.g., Gaussian, Gamma) can be fitted. Subtracting the EQ-5D questionnaire score from 1 produces a right skewed variable for which log links may be appropriate.

It could also be debated that the EQ-5D questionnaire is not actually censored at 1, but this represents a true maximum and should be treated as such. If there are a sufficient number of maximum scores and these are not considered censored, then a 2PM similar to the one outlined by Mullahy [33] may be a viable candidate. Traditionally, 2PM have been used to model health care utilization/cost data where a significant proportion of patients incur no resources/costs. For expected cost estimation, a 2PM would consist of a regression in which typically the first part of the equation is the probability of the cost exceeding 0 estimated over all subjects, usually using a probit/logit specification, and the second part is the expected cost estimated only over those subjects with strictly positive costs, that is,

$$E[y|x] = \Pr(y > 0|x)E[y|y > 0, x]$$

where E is the expected value, $y \geq 0$ the cost, and x covariates.

For our purposes, cost is replaced by EQ-5D questionnaire values, so that the regression now estimates the expected EQ-5D questionnaire outcome. Because there is a predominance of EQ-5D questionnaire values at the maximum of 1, a pattern opposite to that typically seen for cost data (i.e., tendency to 0 values) emerges. Hence, for the utility modeling, we calculated a “Reverse EQ-5D questionnaire score” (i.e., $1 - \text{EQ-5D questionnaire score}$) and then adopted a number of steps to handle these data statistically in a 2PM. This consisted of running a logistic regression to estimate the probability of Reverse EQ-5D questionnaire score being more than 0. In those patients with a positive Reverse EQ-5D questionnaire value, a GLM was used with Reverse EQ-5D questionnaire score as the dependent variable. The coefficients from this analysis were then used to estimate the expected value of the Reverse EQ-5D questionnaire for each patient in the sample, conditional upon Reverse EQ-5D questionnaire score being more than 0. The final step was to transform the resulting value back to provide a predicted EQ-5D questionnaire value. The equations to calculate the expected value of the Reverse EQ-5D questionnaire are presented in a technical appendix to this article in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2012.12.005>. To our knowledge, using such a 2PM applied to $1 - \text{EQ-5D}$ has not been attempted before, and so represents a novel approach in the utility mapping field. Specifically, by entering the Reverse EQ-5D questionnaire into a GLM model, a right skewed variable was created, allowing the log link function to apply its greatest potential benefit. The strength of the approach was explored separately for the idiopathic and neurogenic patient samples.

Explanatory/Predictor Variables Used and Selection and Transformation Technique

The same modeling techniques were applied to both the idiopathic and neurogenic samples for this modeling exercise. Two candidate predictor sets were used, both of which included age. In the first, only the I-QOL total score accompanied age. In the second, the total score was replaced by its three constituent

domains of avoidance and limiting Behaviors, psychosocial impacts, and social embarrassment. All analyses were conducted by using STATA Version 10.1. Multivariable fractional polynomial (MFP) models, as implemented in the Stata package, were used to guide the selection within each predictor set by subgroup. This package searches for the best polynomial terms for each predictor and also performs a backwards stepwise regression once these are found to eliminate unnecessary variables ($P > 0.05$ used as cutoff). The process is repeated to convergence. Thus, the package is intended to produce the best (potentially) nonlinear model eliminating insignificant predictors [34]. While this method may exclude potentially influential variables, it helps to ensure the best choice of functional form for the stronger predictors [35]. All nonlinear transformations were checked to ensure that the effects were plausible through the whole range (e.g., checking that at no point did an increasing adult age predict higher EQ-5D questionnaire scores). Thus, via the MFP and subsequent plausibility checks, it was possible for the same methodology to produce different selections and transformations of predictors between the neurogenic and idiopathic subgroups.

Model performance was assessed by 10-fold cross-validation on the basis of model accuracy measures, mean absolute error (MAE), and root mean squared error (RMSE), which consists of splitting the sample into 10 equally sized groups via a simple randomization process and performing statistical analyses on 9 of the 10 groups to generate model parameter estimates. These parameters are applied to the “left out tenth” group to generate predictions. The MAE and RMSE are then calculated by comparing these predictions to actual values. This process is repeated 10 times, so that all groups are “left out” once.

In the instances in which there was disagreement between the accuracy measures for the choice of the best model, RMSE was selected owing to the fact that it is the most commonly applied “loss function” criteria applied in decision-making models [36]. The squared correlation coefficients between the fitted values for the best model and the actual values of the EQ-5D questionnaire (US tariff) were also shown as a measure of goodness of fit. In the OLS regressions, this value equals the R^2 statistic, or proportion of variance explained. For univariate comparisons of descriptive mean statistics, t tests assuming unequal variances were used. Potential multicollinearity problems between predictors were assessed by variance inflation factor (VIF) scores and the condition number. VIF scores above 5 and a condition number greater than 20 are considered indicative of colinearity problems [36,37]. Bootstrapping techniques (bias corrected with 1000 replications) were utilized to obtain 95% confidence intervals around the best 2PM predictions.

Results

Sample Characteristics and Descriptive Statistics

The number of patients who met the inclusion criteria and completed the necessary questionnaires was 2605, of which 254 were of neurogenic etiology and 2351 were of idiopathic etiology.

Basic demographics and condition status information are presented in Table 1. In both the United States and Europe, idiopathic patients scored higher on health status measures (EQ-5D questionnaire, I-QOL total score, and I-QOL domain scores) than did neurogenic patients (all $P < 0.001$). When US and UK preference value tariffs were applied to the EQ-5D questionnaire scores, the results were very similar; the choice of best models and predictors was the same across tariffs. For consistency, the US tariff was therefore applied for all subsequent analyses.

Table 1 – Demographics and condition status descriptive statistics from Adelphi OAB Disease Specific Programme 2010.

	Number	Percentage within heading group	Mean age (y)	Mean EQ-5D questionnaire score (US tariff)	Mean total I-QOL score	I-QOL domain 1 score	I-QOL domain 2 score	I-QOL domain 3 score
Sample	2605	100.0	58.7	0.85	62.8	60.2	67.3	59.0
Neurogenic	254	9.8	48.7	0.71	53.3	53.4	56.0	48.2
USA	67	26.4	52.5	0.71	59.5	59.7	62.6	53.9
Europe	187	73.6	47.4	0.72	51.0	51.1	53.7	46.1
Male	80	31.5	48.0	0.73	49.5	50.4	50.6	46.3
Female	174	68.5	49.1	0.71	55.0	54.7	58.5	49.0
Idiopathic	2351	90.2	59.8	0.87	63.9	60.9	68.6	60.1
USA	586	24.9	57.5	0.90	71.4	67.6	76.5	68.3
Europe	1765	75.1	60.5	0.86	61.4	58.7	65.9	57.4
Male	334	14.2	61.4	0.88	67.3	65.5	70.4	64.7
Female	2017	85.8	59.5	0.87	63.3	60.1	68.2	59.4

EQ-5D, EuroQol five-dimensional; I-QOL, Incontinence-specific Quality of Life Questionnaire; I-QOL domain 1, avoidance and limiting behavior; I-QOL domain 2, psychosocial impacts; I-QOL domain 3, social embarrassment; OAB, overactive bladder.

The EQ-5D questionnaire scores of both idiopathic and neurogenic patients were subject to ceiling effects, with a more pronounced effect in the idiopathic sample, as shown by 46% of the patients scoring a perfect 1 compared with 20% for neurogenic patients (Fig. 1). In the idiopathic sample, one patient (0.04%) scored below 0 and 98% scored above 0.4. In the neurogenic sample, two patients scored below 0 (0.8%) and 87% scored above 0.4. The high proportion of patients who scored a perfect 1 suggests that a 2PM may be most appropriate to use in the idiopathic sample.

Pearson correlation coefficients were examined between I-QOL and EQ-5D questionnaire scores for both groups. Correlations between I-QOL domains and the EQ-5D questionnaire were stronger and had a smaller range in the idiopathic sample ($0.36 \leq r \leq 0.38$, $P < 0.001$) than in the neurogenic sample ($0.22 \leq r \leq 0.35$, $P < 0.001$). Increasing age was statistically cor-

related with lower EQ-5D questionnaire scores in the idiopathic sample ($r = -0.25$, $P < 0.001$), but such a link was absent in the neurogenic sample, suggesting that the severity of the underlying neurogenic condition outweighs any age-related decrement in HRQOL ($r = -0.06$, $P > 0.05$). It has been shown that valuations of health fall with increasing age, with older individuals reporting more problems on all dimensions [37]. Because age did not present as a significant covariate in the neurogenic algorithm to estimate the EQ-5D questionnaire values, this was not incorporated into the utility estimate, thus making use of the estimated utility in economic models potentially more difficult. Potential collinearity between model predictors was tested by using both VIF and condition numbers, which fell within acceptable limits (all VIF scores < 5 and condition number < 20 .) Thus, it was concluded that collinearity was not an issue.

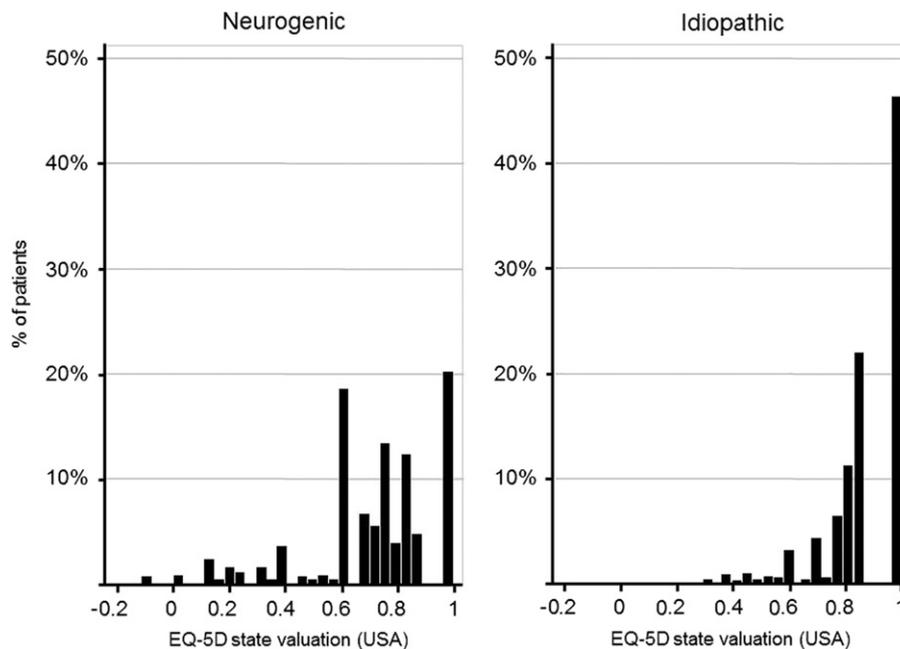


Fig. 1 – Distribution of EQ-5D scores for neurogenic and idiopathic patients.

Table 2 – RMSEs and MAEs for idiopathic models over 10 rotated validation subgroups.

Technique*	Errors	I-QOL total score + age	MFP – linear I-QOL total score only	I-QOL 3 domains + age	MFP – linear I-QOL SE domain only
OLS	RMSE	0.1384	0.1372	0.1384	0.1370*
	MAE	0.1039	0.1036	0.1042	0.1045
CLAD	RMSE	0.1417	0.1415	0.1425	0.1421
	MAE	0.1001†	0.1001†	0.1006	0.1005
Tobit	RMSE	0.1506	0.1509	0.1506	0.1518
	MAE	0.1025	0.1021	0.1026	0.1030
GLM	RMSE	0.1383	0.1372	0.1384	0.1371
	MAE	0.1035	0.1035	0.1038	0.1045
RGLM	RMSE	0.1380	0.1383	0.1379	0.1385
	MAE	0.1069	0.1079	0.1071	0.1086
R2GLM	RMSE	0.1379	0.1370*	0.1379	0.1370*
	MAE	0.1048	0.1047	0.1050	0.1055

CLAD, censored least absolute deviation; GLM, generalized linear model; I-QOL, Incontinence-specific Quality of Life Questionnaire; MAE, mean absolute error; MFP, multivariable fractional polynomial; OLS, ordinary least squares; RGLM, reverse generalized linear model; R2GLM, reverse two-part generalized linear model; RMSE, root mean squared error; SE, social embarrassment.

* Denotes best RMSE result in table.

† Denotes best MAE result in table; bold text denotes best RMSE and MAE result in each column.

Mapping Results

While the same model fitting techniques were used for both the neurogenic and idiopathic samples, the best fitting model was different for each group.

Idiopathic Sample

Table 2 summarizes the various models tested for the idiopathic sample, along with the corresponding mean RMSE and MAE over the 10 rotated validation subgroups. Tested combinations of predictors of interest included age, I-QOL total score, and the I-QOL constituent domains (avoidance and limiting behavior, psychosocial impacts, and social embarrassment).

In the idiopathic OAB sample, the choice of model predictors does not make a substantial difference to the predictive success rates for any one statistical model, because results vary only in the third or fourth decimal. CLAD is superior to Tobit on all measures. Results across all other (noncensoring) statistical methods were very similar. Given that a reverse 2PM produces the best RMSE, further improvements were tested by allowing different nonlinear transformations in the two separate parts of

the model. Accordingly, separate MFP models were run (data not shown), and it was found that different nonlinear transformations in the two parts were appropriate. Slightly better fit results followed (lowest RMSE = 0.1369, lowest MAE = 0.1037 for the R2PM model, using a Gaussian Family and Log link in GLM for the second part of the model).

In the first part of the model, the probability of scoring the EQ-5D questionnaire as less than 1 was modeled by using nonlinear squared transformations of both I-QOL total score and age. Age itself was initially transformed to the number of positive years over 55. For example, an individual aged 65 years would be assigned a value of 10. Such a pre-transformation was necessary to prevent the model from predicting a slight rise in the probability of a perfect score (EQ-5D questionnaire score = 1) as the individual increases in age from twenties to fifties.

For the second part of the model, it was not necessary to apply a preliminary transformation to age as occurred in the first part. The best model used a cubic transformation on I-QOL total score and a square root transformation on age (resulting equations are summarized in the technical appendix). It was possible by combining both equations to show the total effect on predicted EQ-5D questionnaire score for both I-QOL total score and age. These are shown in Figures 2 and 3.

It is clear from Figure 3 that aging has an influence on the EQ-5D questionnaire score only in later life, which explains to some degree why I-QOL total score registered as a more statistically significant predictor than age in both models. The squared correlation between the predictions and the actual EQ-5D questionnaire values (R^2 statistic) for this 2PM is 0.21.

The predictions generate an EQ-5D questionnaire score range of 0.68 to 0.97 (with age set to the mean of 59.8 years) as the I-QOL total score varies between 0 and 100, respectively (see technical appendix in Supplemental Materials found at <http://dx.doi.org/10.1016/j.jval.2012.12.005> for actual formula needed to predict EQ-5D questionnaire score <http://dx.doi.org/10.1016/j.jval.2012.12.005> score from I-QOL).

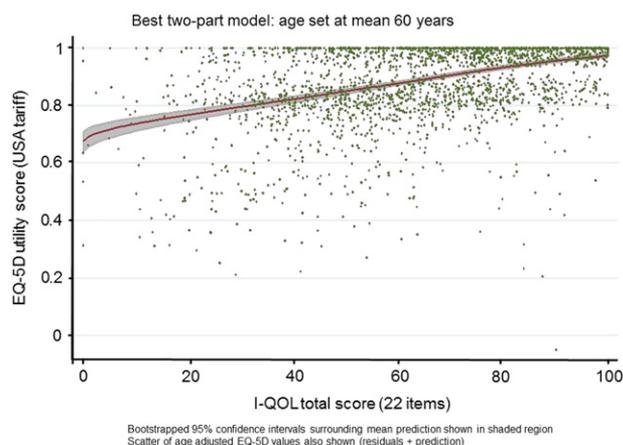


Fig. 2 – Idiopathic sample - predicted EQ-5D scores by I-QOL total score.

Neurogenic Sample

Table 3 summarizes the various models tested for the neurogenic sample, along with the corresponding RMSE and MAE. Similarly

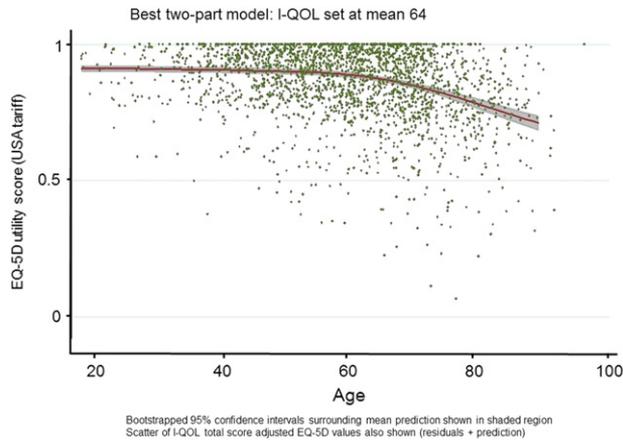


Fig. 3 – Idiopathic sample: predicted EQ-5D scores by age.

to the idiopathic sample, predictors of interest included combinations of age, IQOL total score, and the IQOL constituent domains.

For the neurogenic sample, age is omitted from both MFP models, indicating that it is not a statistically significant predictor ($P = 0.80$ and $P = 0.70$ in the two models for I-QOL total score and I-QOL social embarrassment domain score, respectively). Notable is that the MFP routines do not find any nonlinear transformations in preference to linear. The reverse GLM using a Normal Family and Log link is the best model for all predictors as judged on RMSE criteria. It also performs relatively well on MAE criteria and is considered the best model in conjunction with using I-QOL social embarrassment domain score as the sole predictor alongside a constant. The log link itself ensures a nonlinear effect, which is shown in Figure 4.

The predictions generate a limited EQ-5D questionnaire score range of 0.47 to 0.87 as the I-QOL social embarrassment domain score varies between 0 and 100, respectively (see technical appendix in Supplemental Materials found at: <http://dx.doi.org/10.1016/j.jval.2012.12.005> for actual formula needed to predict the EQ-5D questionnaire score from I-QOL). The squared correlation

between the predictions and the actual EQ-5D questionnaire values for this best model (R^2 statistic) is 0.13.

Actual versus Predicted EQ-5D Questionnaire Scores

The actual and predicted EQ-5D questionnaire scores, along with I-QOL scores for both the idiopathic and neurogenic samples, are summarized in Table 4. For both groups, the mean actual score is very close to EQ-5D questionnaire scores predicted by each respective model; however, the largest discrepancy is noted in the neurogenic incontinent subgroup (0.698 and 0.728 for actual and predicted EQ-5D questionnaire scores, respectively).

Sensitivity Analysis

Additional analyses were conducted to validate these findings. The same mapping exercise described above was performed independently among four subgroups (USA idiopathic, USA neurogenic, Europe idiopathic, and Europe neurogenic) to ensure that the resulting idiopathic and neurogenic models remained robust and that no fundamental difference across subgroups was noted.

In addition, analysis was conducted by mapping I-QOL onto each of the EQ-5D questionnaire domains separately. Generalized ordered logit regressions were fitted (the proportional odds/parallel lines assumption was tested and if not rejected, then a parallel lines model was fitted instead of the generalized ordered logit) by using the I-QOL domains and age, including nonlinear terms (using MFP as used in the previous analyses), and the backward stepwise approach was used (removing variables with $P > 0.05$). Out of sample predictions were used (from 10-fold cross-validation) to generate a predicted EQ-5D questionnaire score and hence MAE and RMSE values. These results (summarized in Table 1 of the technical appendix) suggest that this method does not outperform the best solution as previously generated in the base-case analysis.

Discussion

The aim of the mapping study was to develop an algorithm to map the I-QOL, a urinary incontinence-specific instrument, onto

Table 3 – RMSEs and MAEs for neurogenic models over 10 rotated validation subgroups.

Technique*	Errors	I-QOL total score + age	MFP – linear I-QOL total score only	I-QOL 3 domains + age	MFP – linear I-QOL social embarrassment domain only
OLS	RMSE	0.2269	0.2263	0.2898	0.2233
	MAE	0.1679	0.1680	0.1686	0.1667
CLAD	RMSE	0.2307	0.2285	0.3155	0.2263
	MAE	0.1692	0.1685	0.1718	0.1685
Tobit	RMS	0.2303	0.2293	0.3035	0.2265
	MAE	0.1688	0.1693	0.1703	0.1685
GLM	RMSE	0.2274	0.2268	0.2962	0.2236
	MAE	0.1689	0.1690	0.1746	0.1673
RGLM	RMSE	0.2255	0.2251	0.2857	0.2221*
	MAE	0.1672	0.1674	0.1677	0.1662†
R2GLM	RMSE	0.2271	0.2263	0.2903	0.2233
	MAE	0.1681	0.1680	0.1687	0.1666

CLAD, censored least absolute deviation; GLM, generalized linear model; I-QOL, Incontinence-specific Quality of Life Questionnaire; MAE, mean absolute error; MFP, multivariable fractional polynomial; OLS, ordinary least squares; RGLM, reverse generalized linear model; R2GLM, reverse two-part generalized linear model; RMSE, root mean squared error.

* Denotes best RMSE result in table.

† Denotes best MAE result in table; bold text denotes best RMSE and MAE result in each column.

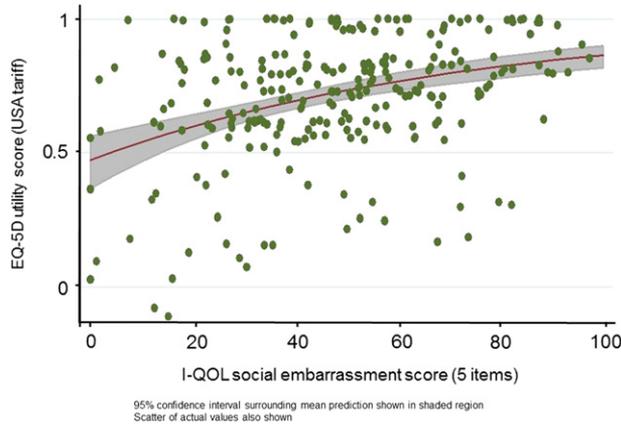


Fig. 4 – Neurogenic sample: predicted average EQ-5D score via best model I-QOL social embarrassment score only predictor.

the generic EQ-5D questionnaire measure, using data from both idiopathic and neurogenic patient samples. The ultimate aim was to identify the statistical models with best predictive power to enable future estimation of utility values based on I-QOL data collected in clinical studies of idiopathic or neurogenic QAB syndrome.

There has been a growth in mapping studies in recent years, with several recently published in *Value in Health* that report mapping a disease-specific instrument to the EQ-5D questionnaire and covering areas such as cancer [38–40], osteoarthritis [41], and chronic obstructive pulmonary disease [42]. A study has also been published that generated a set of preference values for OAB on the basis of a survey of the UK general population by using the five-dimensional health classification system (OAB-5D), derived from the validated Overactive Bladder Questionnaire [43]. Deriving such a condition-specific preference-based measure from a disease-specific instrument is a lengthy process, requiring a direct utility elicitation survey to be designed to collect the relevant data. A similar single index preference measure has not been generated for the I-QOL. Given that there are available data for both the disease-specific I-QOL and the generic EQ-5D questionnaire from an observational study, mapping provides a more rapid solution to generating utilities that are related to the condition-specific instrument. In this research for the mapping exercise, we used the available Adelphi OAB DSP, which contains I-QOL and EQ-5D questionnaire data for more than 2600 idiopathic and neurogenic OAB patients. This is the first time that a mapping algorithm has been applied to generate EQ-5D

questionnaire utilities on the basis of the I-QOL measure. A descriptive analysis of the OAB data set demonstrated that there is a relationship between both higher I-QOL and EQ-5D questionnaire scores and being dry (i.e., no incontinence episodes) versus wet.

As in all utility mapping exercises, the primary first step is to identify the model that best fits the data. By using the OAB DSP, we found that for the idiopathic patient sample, the best fitting model included the I-QOL total score and age (both modeled nonlinearly). For the neurogenic patient sample, the best model contained the I-QOL social embarrassment domain score; this was modeled linearly only. A key aspect of the robustness of mapping is the predictive strength of the best fitting model. From our analysis, the reverse GLM using a Normal Family and Log link provided the best predictive strength models for all predictors as judged on RMSE criteria, for idiopathic and neurogenic populations, respectively. The relatively better fitting results were in the idiopathic patient sample ($n = 2351$; MAE = 0.10; RMSE = 0.14) than in the neurogenic sample ($n = 254$; MAE = 0.17; RMSE = 0.22), with this difference likely to be related to the much smaller sample size for the latter. The range of EQ-5D questionnaire score predictions generated by the models could be considered narrow at 0.68 to 0.97 and 0.47 to 0.87 for the idiopathic and neurogenic patient samples respectively. It seems reasonable, however, that the bounds are generally reflective of the range of HRQOL impact associated with OAB, with, as might be expected, a greater range seen in patients with neurogenic relative to idiopathic etiology.

The exploration and use of a two-part reverse model in this study's mapping may represent a departure from most previous mapping exercises that focus on OLS models and alternatives such as Tobit or CLAD models. The 2PM has previously been used for cost estimation, but it has been adapted here by using a reverse function to address the bounding of utilities to 1 rather than 0 as is the case with costs.

While the number of mapping studies has increased, they are still in their relative infancy and, as with previously published mapping evaluations, there are limitations with our analysis. In particular, the performance of our mapping exercise is at the lower end of those reviewed by Brazier et al. [44], which found RMSE values ranging from 0.084 to 0.2 for a range of conditions including OAB, rheumatoid arthritis, Crohn's disease, stable angina, and asthma. The mapping model between the six-dimensional health state short form (derived from short-form 36 health survey) and the Overactive Bladder Questionnaire covered in the Brazier et al. [44] review was stated to represent one of the poorer fitting models in the review. Hence, this is clearly a condition in which to date it has been difficult to fit mapping models. It should be noted that for conditions such as OAB that are targeted to one area of the body, mapping to a

Table 4 – Actual versus predicted EQ-5D questionnaire scores (US tariff).

	Idiopathic		Neurogenic	
	Wet	Dry	Wet	Dry
n	374	198	58	8
I-QOL total summary score, mean \pm SD	65.7 \pm 21.3	81.8 \pm 16.5	58.2 \pm 24.4	69.9 \pm 17.7
Actual EQ-5D questionnaire score, mean \pm SD	0.885 \pm 0.138	0.928 \pm 0.117	0.698 \pm 0.239	0.753 \pm 0.309
Predicted EQ-5D questionnaire score, mean \pm SD	0.874 \pm 0.071	0.925 \pm 0.053	0.728 \pm 0.098	0.761 \pm 0.082

Note. Results based on data from US survey participants. Stratification into “wet” or “dry” was based on patients' responses to the question asking them: “Are there times when you experience leakages, however small the amount?” Patients who answered “Yes” to this question were classified as “wet,” while those answering “No” to this question were classified as dry.
EQ-5D, EuroQol five-dimensional.

general measure is less sensitive and may leave a substantial degree of the disease-specific impact unaccounted for. Therefore, even with the additional modeling methods, such as the two-part reverse model, the lack of strong predictive ability of the mapping may be indicative of a relatively low degree of conceptual relationship between the EQ-5D questionnaire and the I-QOL. Lack of conceptual overlap between the chosen instruments is a general limitation of mapping studies, and one that may constrain future mapping exercises in the field of OAB with the current HRQOL instruments available. Given this, an interesting future approach may be to explore the strength of the relationship between the I-QOL and the recently developed preference-based measure, the five-dimensional OAB [43].

In addition, the estimation of an algorithm for the neurogenic OAB patients was constrained by the sample size. In contrast, the idiopathic model was based on a large patient sample. There is a need to expand the sample size in neurogenic patients, which should at least improve the predictive strength of the best fitting model for this patient group closer to that estimated for the idiopathic patients. In addition, the effectiveness of the 10-fold cross-validation technique to calculate goodness of fit was limited in the smaller neurogenic sample.

Another issue is the appropriate EQ-5D questionnaire preference tariff to use for the mapping exercise. For this article, the analysis was performed by using the US tariff. Other options are to use the UK tariff (as the longest established tariff), or apply the US tariff to US patients and the UK tariff to European patients. This was explored in this study, but the results remained the same irrespective of which tariff was used. This study used the current three-level EQ-5D questionnaire. In future, this may be replaced in clinical studies by the five-level EQ-5D questionnaire, which should reduce the ceiling effect problems associated with the three-level instrument. Ceiling effects were observed in this study with a large proportion of patients in the data set recording an EQ-5D questionnaire score of 1. This may have reduced the ability of the instrument to transfer improvements in HRQOL, associated with reductions in urinary incontinence, into an estimated utility improvement.

Although a convenient sample of patients was stipulated by the Adelphi OAB DSP, with physicians instructed to record full details for the next 10 consecutive consulting patients who met the eligibility criteria, these data were collected in real-world clinical practice with no specific interventions imposed by protocol. In addition, the study was conducted at multiple centers throughout the United States and Europe. These results may therefore be considered generalizable, within limits, to the broader idiopathic and neurogenic OAB populations.

A large number of countries, including the United Kingdom, the Netherlands, Sweden, Poland, Australia, and Canada, use cost per quality-adjusted life-year as the criterion for aiding drug reimbursement and access decisions. We have produced mapping models that can in principle be applied to clinical trials in idiopathic and neurogenic OAB that contain the I-QOL to generate utilities based on the EQ-5D questionnaire. In the United Kingdom, methods guidance for technology appraisal issued by the National Institute for Health and Clinical Excellence (NICE) and other health technology assessment (HTA) bodies such as the Scottish Medicines Consortium specify mapping approaches for cost-utility analyses as an acceptable alternative option to using direct data from generic instruments, in particular the EQ-5D questionnaire, in trials. A review of utility methods used in NICE technology appraisals found that 27% (19 of 46 appraisals) used some form of mapping approach to generate utilities for use in cost-utility analyses in the years 2004 to 2008, with 14 of these involving a mapping to the EQ-5D questionnaire. An update of this review found that mapping to the EQ-5D questionnaire had been used in a further two appraisals, although there was a lack

of reporting of the statistical properties of the mapping algorithms [45]. To improve the quality and transparency of mapping studies used in technology appraisals, the NICE Decision Support Unit has recently provided a set of recommendations on mapping methods that includes the full description of the data set used to estimate the mapping regression including both the range of EQ-5D questionnaire values and graphical plots showing the distribution of EQ-5D questionnaire data, and statistical properties of the mapping algorithms should be clearly described, including reporting of the RMSE or MSE [46]. We believe that our analysis meets the core recommendations for the presentation of the methods and results contained in this report. Given the limitations in the conceptual relationship between the two instruments and the predictive strength of the association outlined above, especially for the neurogenic patient group, however, for the data are to be used for HTA, it is important that sufficient sensitivity analysis is performed on the utilities generated.

In conclusion, there is a need for pragmatism in the use of mapped utilities in HTA for decision making. Therefore, in the absence of a viable EQ-5D questionnaire source, results from the I-QOL to EQ-5D questionnaire algorithms produced here are of value in providing potential current utility data for idiopathic and neurogenic OAB economic models supplied as part of NICE, Scottish Medicines Consortium, and other HTA body assessments. For future HTA and other decision-making purposes, however, there is an ongoing need for further investigation and data analysis in an attempt to improve the performance of mapping studies in this field.

Acknowledgments

Some of these results were presented at a podium presentation at the 14th European Congress of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) in Madrid, Spain, in November 2011. The authors are grateful to the audience for questions and comments on the research, in particular in the area of sensitivity analysis. The authors acknowledge the contribution of Gary Milligan, Head of Statistics at Adelphi Real World, for running these additional analyses.

Source of financial support: Financial support for this study was provided by a contract with Allergan, Inc., Irvine, CA, USA. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. S.K. and P.A. contributed to the conception and design of the Disease Specific Programme. S.K. conducted the statistical analyses. All the authors contributed to the interpretation of data; drafting the article, revising it critically for intellectual content, and final approval of the version to be published. None of the authors has any competing interests. All authors were employees of either Adelphi or Allergan at the time this research was conducted or received compensation from Allergan for consulting services provided to them to conduct this research. S.K. was an employee of Adelphi Real World at the time of the study but has since left the company. K.T. is an independent health economics consultant. D.C., an employee of Allergan, Inc., at the time of the study, has since left the company but continues to act as an independent consultant for Allergan. All other authors are employees of Adelphi Real World or Allergan, Inc.

Supplemental Materials

Supplemental material cited in this article is available online.

REFERENCES

- [1] Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
- [2] Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986;5:1–30.
- [3] Szende A, Schaefer C. A taxonomy of health utility assessment methods and the role for uncertainty analysis. *Eur J Health Econ* 2006;7:147–51.
- [4] McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales and Questionnaires* (2nd ed.). New York: Oxford University Press, 1996.
- [5] Hurst NP, Kind P, Ruta D, et al. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 1997;36:551–9.
- [6] Brazier J, Jones N, Kind P. Testing the validity of the EuroQol and comparing it with the SF-36 health survey questionnaire. *Qual Life Res* 1993;2:169–80.
- [7] Kind P, Hardman G, Macran S. UK population norms for EQ-5D. York Centre for Health Economics, Discussion Paper. University of York 1999.
- [8] Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care* 2005;43:203–20.
- [9] Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011;20:1727–36.
- [10] Abrams P, Cardozo L, Fall M, et al. The standardisation of lower urinary tract function: report from the Standardisation Sub-committee of the International Continence Society. *Neurourol Urodyn* 2002;21:167–78.
- [11] Milsom P, Abrams P, Cardozo L, et al. How widespread are the symptoms of an overactive bladder and how are they managed? A population-based prevalence study. *BJU Int* 2001;87:760–6.
- [12] Stewart WF, Van Rooyen JB, Cundiff GW, et al. Prevalence and burden of overactive bladder in the United States. *World J Urol* 2003;20:327–36.
- [13] Irwin DE, Milsom I, Hunskaar S, et al. Population-based survey of urinary incontinence, overactive bladder, and other lower urinary tract symptoms in five countries: results of the EPIC study. *Eur Urol* 2006;50:1306–14.
- [14] Herschorn S, Gajewski J, Schultz J, Corcos J. A population-based study of urinary symptoms and incontinence: the Canadian Urinary Bladder Survey. *BJU Int* 2007;101:52–8.
- [15] Abrams P, Kelleher CJ, Kerr LA, Rogers RG. Overactive bladder significantly affects quality of life. *Am J Manag Care* 2000;6:S580–90.
- [16] Tubaro A. Defining overactive bladder: epidemiology and burden of disease. *Urology* 2004;64(Suppl):S2–6.
- [17] Coyne KS, Payne C, Bhattacharyya SK, et al. The impact of urinary urgency and frequency on health-related quality of life in overactive bladder: results from a national survey. *Value Health* 2004;7:455–63.
- [18] Irwin D, Milsom I, Kopp Z, et al. Impact of overactive bladder symptoms on employment, social interactions and emotional well-being in six European countries. *BJU Int* 2006;97:96–100.
- [19] Zorn BH, Montgomery H, Pieper K, et al. Urinary incontinence and depression. *J Urol* 1999;162:82–4.
- [20] Norton PA, Zinner NR, Yalcin I, Bump RC. Duloxetine versus placebo in the treatment of stress urinary incontinence. *Am J Obstet Gynaecol* 2002;187:40–8.
- [21] Dmochowski RR, Miklos JR, Norton PA, et al. Duloxetine versus placebo for the treatment of North American women with stress urinary incontinence. *J Urol* 2003;170:1259–63.
- [22] Millard RJ, Moore K, Rencken R, et al. Duloxetine vs placebo in the treatment of stress urinary incontinence: a four-continent randomized clinical trial. *BJU Int* 2004;93:311–8.
- [23] van Kerrebroeck P, Abrams P, Lange R, et al. Duloxetine versus placebo in the treatment of European and Canadian women with stress urinary incontinence. *BJOG* 2004;111:249–57.
- [24] Patrick DL, Martin ML, Bushnell DM. *The I-QOL: A Quality-of-Life Instrument Specific to Persons with Urinary Incontinence: User's Manual and Scoring Diskette for United States Version*. Seattle, WA: Health Research Associates, Inc, 1999.
- [25] Wagner TH, Patrick DL, Bavendam TG, et al. Quality of life of persons with urinary incontinence: development of a new measure. *Urology* 1996;47:67–71.
- [26] Patrick DL, Martin ML, Bushnell DM, et al. Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL). *Urology* 1999;53:71–6.
- [27] Patrick DL, Martin ML, Bushnell DM, et al. Cultural adaptation of a quality-of-life measure for urinary incontinence. *Eur Urol* 1999;36:427–35.
- [28] Bushnell DM, Martin ML, Summers KH, et al. Quality of life of women with urinary incontinence: cross-cultural performance of 15 language versions of the I-QOL. *Qual Life Res* 2005;14:1901–13.
- [29] Schurch B, Denys P, Kozma CM, et al. Reliability and validity of the Incontinence Quality of Life questionnaire in patients with neurogenic urinary incontinence. *Arch Phys Med Rehabil* 2007;88:646–52.
- [30] Anderson P, Benford M, Harris N, et al. Real-world physician and patient behaviour across countries: disease-specific programmes – a means to understand. *Curr Med Res Opin* 2008;24:3063–72.
- [31] National Institute for Health and Clinical Excellence. Submission to the Kennedy Review April 3, 2009. Available from: <http://www.nice.org.uk/media/B4F/7F/KennedyPHE.pdf>. [Accessed October 2, 2012].
- [32] Sullivan PW, Ghushchyan V. Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. *Med Decis Making* 2006;26:401–9.
- [33] Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *J Health Econ* 1998;17:247–81.
- [34] Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs. *Comput Stat Data Anal* 2006;50:3464–85.
- [35] Royston P, Sauerbrei W. *Multivariable Model-Building*. New York: Wiley, 2008.
- [36] Judge GG, Hill RC, Griffiths WE. *Introduction to the Theory and Practice of Econometrics* (2nd ed.). New York: Wiley, 1988.
- [37] Greene W. *Econometric Analysis* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [38] Kontodimopoulos N, Aletras VH, Paliouras D, Niakas D. Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. *Value Health* 2009;12:1151–7.
- [39] Crott R, Briggs A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. *Eur J Health Econ* 2010;11:427–34.
- [40] Pickard AS, Ray S, Ganguli A, Cella D. Comparison of FACT- and EQ-5D-based utility scores in cancer. *Value Health* 2012;15:305–11.
- [41] Xie F, Pullenayegum EM, Li SC, et al. Use of a disease-specific instrument in economic evaluations: mapping WOMAC onto the EQ-5D utility index. *Value Health* 2010;13:873–8.
- [42] Starkie HJ, Briggs AH, Chambers MG, Jones P. Predicting EQ-5D values using the SGRQ. *Value Health* 2011;14:354–60.
- [43] Yang Y, Brazier J, Tsuchiya A, Coyne K. Estimating a preference-based single index from the Overactive Bladder Questionnaire. *Value Health* 2009;12:159–66.
- [44] Brazier J, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ* 2010;11:215–25.
- [45] Tosh J, Longworth L, George E. Utility values in NICE technology appraisals. *Value Health* 2011;14:102–9.
- [46] Longworth L, Rowen DL. NICE DSU Technical Support Document 10: the use of mapping methods to estimate health state utility values. 2011. Available from: <http://www.nicesdu.or.uk>. [Accessed March 19, 2012].