

Mapping from Disease-Specific Measures to Utility: An Analysis of the Relationships between the Inflammatory Bowel Disease Questionnaire and Crohn's Disease Activity Index in Crohn's Disease and Measures of Utility

Martin J. Buxton, BA,¹ Loretto A. Lacey, PhD,² Brian G. Feagan, MD,³ Timothy Niecko, MS,⁴ David W. Miller, PhD,⁵ Raymond J. Townsend, PharmD⁵

¹Health Economics Research Group, Brunel University, UK; ²Lacey Solutions, Dublin, Republic of Ireland; ³Robarts Research Institute, University of Western Ontario, London, ON, Canada; ⁴Niecko Health Economics, Escondido, CA, USA; ⁵Elan Pharmaceuticals Inc., San Diego, CA, USA

ABSTRACT

Objectives: To examine the relationship between the Inflammatory Bowel Disease Questionnaire (IBDQ), Crohn's Disease Activity Index (CDAI) and measures of utility (EQ-5D and the SF-6D indexes), and to estimate algorithms to map the two utility values from IBDQ and CDAI scores.

Methods: A large data set from clinical trials in Crohn's disease provided contemporaneous patient responses to all four questionnaires. Paired observations from multiple time-points were analyzed. We calculated mean utility scores by IBDQ and CDAI score deciles; Spearman correlation coefficients for paired observations between IBDQ and EQ-5D ($n = 3320$) and IBDQ and SF-6D ($n = 3230$), and explored regression models using maximum likelihood estimation. The IBDQ/SF-6D model was validated against paired observations from an independent data set.

Results: The IBDQ decile analysis demonstrated a consistent positive relationship with both utility indexes. Correlations

between the IBDQ and both the EQ-5D and SF-6D were statistically significant ($P < 0.0001$), with correlation coefficients of 0.76 and 0.85, respectively. A simple linear model between EQ-5D and IBDQ explained 45% of the variance. The residuals plot for the IBDQ/SF-6D model suggested some nonlinearity and a nonlinear model explained 69% of the variance. In the validation analysis, no statistically significant difference was observed between the mean observed SF-6D and the SF-6D scores estimated using the IBDQ/SF-6D regression model.

Conclusions: Given the strength, consistency, and predictable characteristics of the relationships, the algorithms appear to provide valuable and valid methods to estimate utilities from IBDQ scores (but not CDAI) in trials of Crohn's disease patients that have collected IBDQ scores but not utilities.

Keywords: CDAI, Crohn's disease, EQ-5D, IBDQ, SF-6D, utility.

Background

The increasing use of cost-utility analysis, particularly by reimbursement agencies and national advisory bodies such as the National Institute for Health and Clinical Excellence (NICE) in the UK, imposes requirements for additional data from trials. Ideally, all trials would now include a general measure from which a utility value could be directly attributed, such as the EQ-5D [1], SF-6D [2], or Health Utilities Index (HUI) [3]. In the absence of data from such instruments it may be possible to map their values indirectly from a generic health-related quality of life (HRQoL) or a disease-specific instrument included in the trial. A number of such "algorithms" to map utilities from

generic HRQoL instruments have been published [4,5], as have some from disease-specific measures, such as breathlessness and angina severity in cardiac patients [6]. This article presents such an exercise for mapping from instruments applicable to Crohn's disease, to both EQ-5D and SF-6D and to compare the values obtained.

Crohn's disease is a chronic gastrointestinal disorder characterized by relapsing and remitting inflammation of the gastrointestinal tract. Patients experience substantial impairment in HRQoL as a consequence of abdominal pain, diarrhea and fatigue. In Crohn's patients, disease activity is normally measured using the Crohn's Disease Activity Index (CDAI) [7], and the Inflammatory Bowel Disease Questionnaire (IBDQ), a 32-item disease-specific measure, is commonly used to measure HRQoL life in Crohn's disease [8]. The general HRQoL instrument, the SF-36 has also been used to measure HRQoL in Crohn's patients [9].

Address correspondence to: Martin Buxton, Health Economics Research Group, Brunel University, Uxbridge, Middlesex UB8 3PH, UK. E-mail: martin.buxton@brunel.ac.uk
10.1111/j.1524-4733.2007.00171.x

Although there is increasing recognition of the desirability of cost-utility analysis to inform decision-making, there is no general consensus on the best instrument to use. Although the EQ-5D descriptive system has the advantage of simplicity in data collection, concern is often expressed that its three levels on five dimensions are not sufficiently sensitive to discriminate between small but important differences in health states, particularly those close to full health at which level a "ceiling effect" is often observed [10]. The most common alternative is the SF-6D, which draws selectively from items included in the SF-36 questionnaire [2,10]. The SF-6D may be more sensitive than the EQ-5D in groups with mild to moderate health problems but it is criticized as having a "floor effect" at the lower end of the scale (i.e., more severe cases) [10]. In part this reflects that the SF-6D has a narrower valuation range than the EQ-5D.

Konig et al. [11] have previously reported significant correlations between total IBDQ scores and EQ-5D indexes and between CDAI scores and EQ-5D indexes. Nevertheless, their study did not provide details of the regression equations underlying the relationships observed. The primary goals of our analysis were to establish the statistical relationship between IBDQ total scores and EQ-5D and SF-6D based utility values in patients with Crohn's disease and to compare the resulting values. Secondary goals of this study were to establish the statistical relationship between CDAI total scores and the same two utility instruments. If robust statistical relationships between the IBDQ (or the CDAI) and the EQ-5D utility instruments were established, mapping algorithms could be used to generate utilities from trials that have collected the IBDQ or CDAI but have not collected utility data.

Methods

Source of Data

Paired contemporaneous observations from patients with moderately to severely active Crohn's disease who participated in the Efficacy of Natalizumab as Active Crohn's Therapy (ENACT)-1 and Evaluation of Natalizumab as Continuous Therapy (ENACT-2) clinical trials were included in the analysis. ENACT-1 and ENACT-2 were multinational, randomized, placebo controlled trials of natalizumab and results have been published elsewhere [12]. Data from patients at multiple time points in the two trials were pooled. Questionnaires had been self-administered during scheduled visits to the clinic. For the confirmatory analysis, paired contemporaneous IBDQ and SF-6D observations were taken from an, as yet unpublished, study in patients with moderately to severely active Crohn's disease (ENCORE).

Instruments

IBDQ. The IBDQ is a disease-specific HRQoL questionnaire containing 32 items, which are grouped into four subscales, including: bowel symptoms (10 items), systemic symptoms (five items), emotional function (12 items) and social function (five items) [8]. Each item is scored on a 7-point Likert scale ranging from 1 (worst) to 7 (best). Four subscale scores and one total score can be calculated, by summing up the respective item scores. Scores range from 32 to 224 with higher scores indicating better quality of life. An absolute change of 16 points in the Total IBDQ score, or 0.5 points per question, has been used to define a minimum clinically important difference [13].

Crohn's disease activity index. The CDAI incorporates eight items and scores typically range from 0 to 600, although negative and higher scores are possible [7]. Higher scores indicate more severe disease activity. The CDAI score is calculated by summing weighted scores for three subjective items (number of liquid or very soft stools, abdominal pain and general well-being) recorded by a diary card during a 1-week period, and four objective items (associated symptoms, taking antidiarrheal such as Loperamide/opiates, abdominal mass, hematocrit and body weight). Patients with scores of <150, 150 to 219, 220 to 450 are considered to be in remission, mild disease and moderate to severe disease, whereas those with scores of >450 have very severe disease [13].

EQ-5D. The EQ-5D descriptive system consists of five questions relating to problems in the domains "mobility," "self-care," "usual activities," "pain/discomfort," and "anxiety/depression" [1,14]. Responses in each domain are divided into three ordinal levels, coded: 1, no problems, 2 moderate problems; and 3 extreme problems. A five-digit number based on the five dimensions of the EQ-5D is therefore generated. For example, 11111 would indicate no problems in any domain whereas 33333 would indicate extreme problems in all domains. Health states defined by the five dimensional descriptive system are converted into EQ-5D index values using the time trade-off (TTO) elicited "tariff" of values from a UK general population sample [15].

SF-6D. The SF-36, from which the SF-6D is derived, is a generic HRQoL questionnaire containing 36 items [16]. These are grouped into eight subscales and one health transition item, including physical functioning, role limitation due to physical problems, pain, general health perception, vitality, social functioning, role limitation due to emotional problems and mental health. Subscales scores range from 0 (worst) to 100 (best).

The SF-6D is a preference-based measure of health, derived from the SF-36, with six dimensions: physical

functioning, role limitations, social functioning, pain, mental health, and vitality [2]. These six dimensions each have between two and six levels. An SF-6D health state is defined by selecting one level from each dimension. All responders to the original SF-36 questionnaire can be assigned a SF-6D provided the 11 items used in the six dimensions of the SF-6D have been completed. Health states defined by the six dimensional descriptive systems are converted into SF-6D index values using a standard gamble (SG) elicited tariff of values from a UK general population sample [2].

Statistical Analysis

All analyses of EQ-5D and SF-6D index values were in relation to the corresponding paired IBDQ or CDAI score, unless otherwise stated. Missing data were not estimated so the actual number of paired comparisons varied in the different comparisons. An initial analysis presented the mean EQ-5D and SF-6D tariff-based utility scores for each decile of respondents ranked in terms of their IBDQ (and CDAI) scores. Correlations were investigated between the paired observations using Spearman correlation coefficients. The regression relationships that best describe each of the relationships were determined using the SAS MIXED procedure. The MIXED procedure allows corrections for repeated observations in individual patients to be incorporated. It uses the estimation method of restricted maximum likelihood (REML), also known as residual maximum likelihood. The MIXED procedure also offers more flexibility in specifying the correlation structures, which are useful in repeated measures and random effects models. In all cases, utility values were used as the dependent variable. Nonlinearities in the relationships were examined by investigating the plot of residuals obtained and non-linear models explored accordingly.

The R^2 and mean absolute percentage error (MAPE) statistics were used to establish “goodness of fit.” The R^2 statistic in a regression model measures the percentage of variation in the dependent variable, which is explained by the model, as a fraction of the total variation in the dependent variable. The R^2 is also the squared correlation between the observed values of y and the predicted values of y generated by the regression equation. The MAPE measures the predictive accuracy of the model. This measure is equal to the mean, over all observations in the subset, of the absolute value of the difference between each observation’s observed and predicted value, divided by that observation’s observed value.

$$\text{MAPE} = \sum_i (|y - y_i| / y_i) / N$$

Where y_i = the observed value of the dependent variable for observations i , N = number of observations in the subset of interest.

Table 1 Mean (SD) value of all observations included in the ENACT-1 and ENACT-2 data set

	CDAI total score	IBDQ total score	EQ-5D index value	SF-6D index value
Combined data set				
N	3894	3425	3672	3945
Mean	184	157	0.70	0.68
SD	109	38	0.25	0.15
Median	172	161	0.72	0.65
Minimum	-32	44	-0.32	0.30
Maximum	654	224	1.00	1.00

CDAI, Crohn’s Disease Activity Index; ENACT, Efficacy of Natalizumab as Active Crohn’s Therapy; IBDQ, Inflammatory Bowel Disease questionnaire.

Because other demographic and clinical variables may influence EQ-5D and SF-6D health states, the importance of the following covariates was also investigated for each of the relationships: age, gender, baseline concomitant medication (ASA compounds, immunosuppressants, and corticosteroids). The importance of previous surgery for IBD and allocated treatment (natalizumab or placebo) was also examined in a separate analysis.

Variance covariance matrices were also generated to enable those wishing to apply the regression models to incorporate the variance of the prediction error obtained from the matrix, which will reflect that the coefficients from the models are estimated rather than known.

All analyses were undertaken using SAS (SAS Institute Inc. Cary, NC, USA).

Confirmatory analysis. In order to establish the generalisability of the models, the preferred IBDQ/SF-6D algorithm, generated using the ENACT-1 and ENACT-2 database, was applied to paired observations of IBDQ and SF-6D from a separate clinical trial of patients with moderately to severely active Crohn’s disease (ENCORE). IBDQ observations were used to estimate SF-6D values. The means for the observed and estimated SF-6D values were compared using a t -test. In addition the MAPE was calculated for actual and estimated SF-6D observations.

Results

Sample Statistics

The mean (SD) values for all observation of CDAI, IBDQ, EQ-5D, and SF-6D collected in the ENACT-1 and ENACT-2 trials are presented in Table 1. Nine hundred five patients provided multiple observations from different time-points (mean number of observations per patient 5.365; minimum 3; maximum 8). The actual number of observations ranged from 3425 to 3945 and for paired comparisons (Table 2) from 3230 to 3640.

Decile Analysis

Figure 1 presents an analysis of the mean values for EQ-5D and SF-6D by deciles of scores for the IBDQ

Table 2 Spearman correlation coefficients for paired observations from the combined ENACT-1 and ENACT-2 data set

	EQ-5D	SF-6D
CDAI		
N (pairs)	3575	3640
Correlation coefficient	-0.62	-0.66
IBDQ		
N (pairs)	3320	3230
Correlation coefficient	0.76	0.85

All correlations are $P < 0.0001$.

CDAI, Crohn's Disease Activity Index; ENACT, Efficacy of Natalizumab as Active Crohn's Therapy; IBDQ, Inflammatory Bowel Disease questionnaire.

and Figure 2 presents the same for deciles of CDAI. The mean scores for both EQ-5D and SF-6D analyzed by decile of IBDQ and CDAI show entirely logical monotonic relationships. For IBDQ deciles, the EQ-5D ranges from 0.369 to 0.949 and the SF-6D ranges from 0.499 to 0.903. For CDAI deciles the EQ-5D ranges from 0.437 to 0.904 and the SF-6D ranges from 0.547 to 0.848.

Correlations

The Spearman correlation coefficients, calculated for paired observations between CDAI and EQ-5D, IBDQ and EQ-5D, CDAI and SF-6D, and IBDQ and SF-6D, are presented in Table 2. All correlations were statistically significant ($P < 0.0001$). The correlations obtained with the combined data indicated the strongest correlation was for IBDQ and SF-6D (0.85). This was followed by IBDQ and EQ-5D (0.76). The relationship between CDAI and SF-6D (-0.66) and CDAI and EQ-5D (-0.62) were weaker although statistically significant.

Simple Models – Unadjusted for Covariates

The R^2 values for the simple models without covariates for the four pairs are presented in Table 3. The R^2 for the IBDQ and SF-6D model was the largest with 64% of the variation explained by the relationship between

IBDQ and SF-6D. The R^2 for the IBDQ and EQ-5D model was the next largest with 45% of the variation explained by this relationship. Again the relationships with CDAI and the two utility measures were weaker.

Adjusted Least Square Regression Models

Alternative models including covariates, accounting for potential nonlinearity and correcting for repeated observations are presented in Tables 4 and 5 for SF-6D/IBDQ and EQ-5D/IBDQ relationships, respectively. Visual inspection of the IBDQ/SF-6D plot and the accompanying residuals plot suggested some nonlinearity in the relationship. Examining the results for different versions of the SF-6D/IBDQ models, it appears that the addition of an additional IBDQ² parameter to account for nonlinearity results in an improvement in the variation explained (i.e., R^2) compared with the simple model described in Table 3. Little additional improvement in fit is achieved with the addition of other covariates (Table 4). Based on the rule of parsimony, the preferred model is the second model of the form:

$$\text{SF-6D} = 0.4968 - 0.0011 \text{ IBDQ} + 0.000014 \text{ IBDQ}^2$$

(see Table 4).

Examining all versions of the EQ-5D/IBDQ models, it appears that the addition of an additional IBDQ² parameter does not result in any significant improvement in the variation explained (i.e., R^2). Based on the law of parsimony, the preferred model is the first model of the form:

$$\text{EQ-5D} = 0.03043 + 0.0043 \text{ IBDQ} \text{ (see Table 5).}$$

Inclusion of previous surgery for IBD or treatment allocation (natalizumab or placebo) in the preferred models had no significant effect on the variance explained. The variance covariance matrices for the models presented in Tables 4 and 5 are available on request.

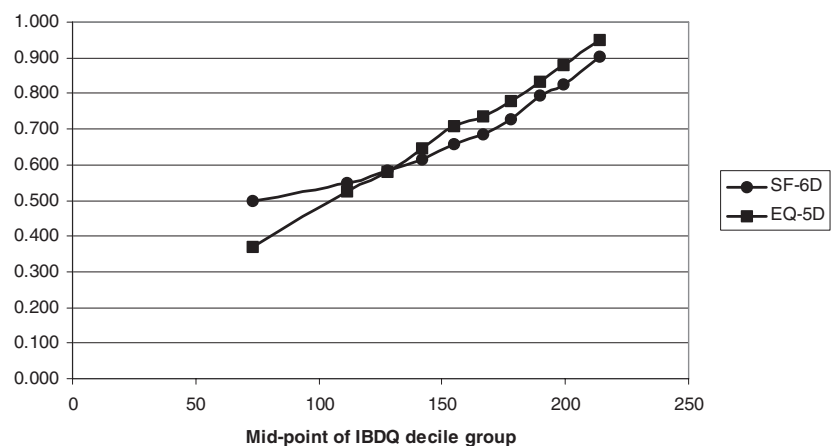


Figure 1 Mean SF-6D and EQ-5D utilities for each Inflammatory Bowel Disease questionnaire (IBDQ) decile subgroup in the combined Efficacy of Natalizumab as Active Crohn's Therapy (ENACT)-1 and ENACT-2 data set.

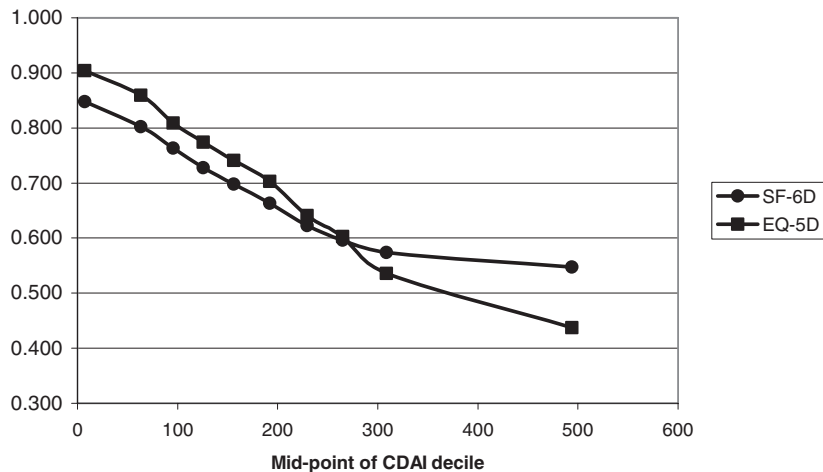


Figure 2 Mean SF-6D and EQ-5D utilities for each Crohn's Disease Activity Index (CDAI) decile subgroup in the combined Efficacy of Natalizumab as Active Crohn's Therapy (ENACT)-1 and ENACT-2 data set.

Confirmatory Analysis

The results of the confirmatory analysis are presented in Table 6. The results show that the preferred IBDQ/SF-6D algorithm predicts actual SF-6D values in an independent Crohn's data set very well. The means for the estimated and observed SF-6D values are not significantly different, and for 78% of the observations the absolute difference between actual and estimated was less than 0.1 (97% were less than 0.2).

Discussion

Although direct collection of data to provide utilities in clinical trials in Crohn's disease remains the method of choice, this analysis suggests that the relationship between the utility measures EQ-5D and SF-6D based on the most commonly used UK valuation sets and the disease-specific HRQoL instrument IBDQ are sufficiently strong to use simple regression equations to estimate either utility value from the IBDQ where that data exist but no utility data have been collected.

Such relationships between the IBDQ and utility measures might be expected. The domain coverage of the IBDQ and EQ-5D overlap on at least three dimensions. For example, "Usual Activities" in the EQ-5D and "Social Function" in the IBDQ both address the impact of illness on ability to perform usual activities (e.g., work, study, etc.); "Anxiety/Depression" in the EQ-5D and the "Emotional Health" dimension in

the IBDQ both address feelings of depression, and the "Pain/Discomfort" dimension in the EQ-5D and the IBDQ's "Bowel symptoms" scale should both capture pain and discomfort associated with illness. Similarly, the content of four out of five domains of the IBDQ appears to overlap with five SF-6D domains: "Role Limitation" (SF-6D) and "Social Functioning" (SF-6D) capture similar impacts as the "Social Function" (IBDQ) scale; the "Pain" (SF-6D) and "Bowel Symptoms" (IBDQ) overlap, as do the Mental Health (SF-6D) and the "Emotional Function" (IBDQ) and the "Vitality" (SF-6D) and "Systemic Symptoms" (SF-6D) both capture impacts of illness on energy levels.

Our observation of slightly higher utility values based on the EQ-5D than the SF-6D at higher IBDQ scale score levels is also consistent with observations made by other authors when comparing the two utility measures in other illnesses [10]. Based on their observations, Brazier et al. suggested that the SF-6D may be better at distinguishing between health states close to full health.

Gregor et al. [13] reported directly elicited utility values from patients with Crohn's disease, using TTO, SG, and visual analog scales (VAS) for both hypothetical states (patients assessment of a described health state and their own current health state). They found that, using direct elicitation, TTO values were significantly higher than SG or VAS. They show a relationship of SG utility and CDAI score broadly similar to

Table 3 Linear models for paired observations from the combined ENACT-1 and ENACT-2 data set

Model type	N pairs	R ²	MAPE (%)	α	SE	B ₁	SE
SF-6D = $\alpha + \beta_1$ IBDQ + error	3230	0.64	9.9	0.2248	0.006	0.0029	0.00004
EQ-5D = $\alpha + \beta_1$ IBDQ + error	3320	0.45	12.2	0.0304	0.0149	0.0043	0.00009
SF-6D = $\alpha + \beta_1$ CDAI + error	3640	0.37	13.6	0.8129	0.0052	-0.00076	0.00002
EQ-5D = $\alpha + \beta_1$ CDAI + error	3575	0.29	8.0	0.9168	0.0082	-0.0012	0.00003

CDAI, Crohn's Disease Activity Index; ENACT, Efficacy of Natalizumab as Active Crohn's Therapy; IBDQ, Inflammatory Bowel Disease questionnaire; MAPE, mean absolute percentage error.

Table 4 Comparison of other models for IBDQ/SF-6D relationship from the combined ENACT-1 and ENACT-2 data set

Model type	N pairs	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	R ²	MAPE (%)
SF-6D = $\alpha + \beta_1$ IBDQ + error	3230	0.2248	0.0029							0.64	9.9
SF-6D = $\alpha + \beta_1$ IBDQ + β_2 IBDQ ² + error	3230	0.4968	-0.0011	0.000014						0.69	9.2
SF-6D = $\alpha + \beta_1$ IBDQ + β_2 Age + β_3 Gender + β_4 Imm + β_5 Ster + error	3230	0.2366	0.0029		ns	0.0088	ns	ns	ns	0.64	9.8
SF-6D = $\alpha + \beta_1$ IBDQ + β_2 IBDQ ² + β_3 Age + β_4 Gender + β_5 Asap + β_6 Imm + β_7 Ster + error	3230	0.508	-0.0012	0.000014	ns	0.0121	ns	ns	ns	0.69	9.2

CDAI, Crohn's Disease Activity Index; ENACT, Efficacy of Natalizumab as Active Crohn's Therapy; IBDQ, Inflammatory Bowel Disease questionnaire; MAPE, mean absolute percentage error; ns, not statistically significant.

Table 5 Comparison of other models for IBDQ/EQ-5D relationship from the combined ENACT-1 and ENACT-2 data set

Model type	N pairs	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	R ²	MAPE (%)
EQ-5D = $\alpha + \beta_1$ IBDQ + error	3320	0.03043	0.0043							0.45	12.2
EQ-5D = $\alpha + \beta_1$ IBDQ + β_2 IBDQ ² + error	3320	ns	0.0052	ns						0.45	12.1
EQ-5D = $\alpha + \beta_1$ IBDQ + β_2 Age + β_3 Gender + β_4 Imm + β_5 Ster + error	3320	ns	0.0043		ns	0.0187	ns	ns	ns	0.46	11.6
EQ-5D = $\alpha + \beta_1$ IBDQ + β_2 IBDQ ² + β_3 Age + β_4 Gender + β_5 Asap + β_6 Imm + β_7 Ster + error	3320	ns	0.0052	ns	ns	ns	ns	ns	ns	0.46	11.6

Age, age at baseline; Asap, ASA compounds taken at baseline; ENACT, Efficacy of Natalizumab as Active Crohn's Therapy; IBDQ, Inflammatory Bowel Disease questionnaire; Imm, immunosuppressant use at baseline; MAPE, mean absolute percentage error; ns, not statistically significant; Ster, corticosteroid use at baseline.

Table 6 Comparison of observed SF-6D and estimated SF-6D values for independent data set predicted using the preferred IBDQ/SF-6D algorithm

	Mean	SD	N
SF-6D actual values (calculated from SF-36)	0.620	0.128	802
SF-6D estimated values (using IBDQ/SF-6D algorithm)	0.621	0.101	802
Mean difference (estimated SF-6D-actual SF-6D)	0.0011	0.084	
t-test	P = 0.72		
MAPE (%)			10.70

IBDQ, Inflammatory Bowel Disease questionnaire; MAPE, mean absolute percentage error.

that in Figure 2. Analyzed according to IBDQ scores our observed utility scores are broadly similar to the hypothetical values elicited by SG. As would be expected, the responsiveness of the estimated values are much higher (7.67 for EQ-5D and 8.00 for SF-6D) reflecting the low standard deviation in the predicted values for the stable patients.

Given that utilities measure something different to either disease activity or disease-specific HRQoL it is difficult properly to test for responsiveness of these utility measures in Crohn's patients. Nevertheless, an indication is provided with a comparison of the mean absolute change in patients whose IBDQ changed by more than 16 points (usually seen as a clinically significant difference) and the standard deviation of the scores of "stable" patients (those who changed by 15 points or less): observed EQ-5D index values produced a responsiveness index score of 0.9 and observed SF-6D a score of 2.19. These figures compare with a value of 1.19 for the UK IBDQ overall, indicating that the EQ-5D is slightly less responsive than the IBDQ although the SF-6D is considerably more responsive [17].

As the relationship between IBDQ and SF-6D and IBDQ and EQ-5D are both sufficiently strong for predictive purposes, it is relevant to establish which of the two utility instruments should be used for quality-adjusted life-year evaluation in Crohn's patients. A number of criteria may inform this debate, including the robustness of the mapping algorithm, the relevance of the instrument content to Crohn's disease, and measurement characteristics. The high proportion of the variance in SF-6D explained by the IBDQ means that the mapping algorithm is unusually strong and should be sufficiently robust for most purposes; while that for the EQ-5D is weaker, it too is good in comparison with some previously published mapping algorithms [6]. There appears to be greater overlap with the SF-6D and the IBDQ than with the EQ-5D and the IBDQ in terms of content as discussed above. As regards their measurement characteristics, each instrument exhibits its known weakness. (Fig. 1) The ceiling effect of the EQ-5D means that it may poorly differentiate among states close to full health: for the

least severe IBDQ decile point, the EQ-5D gives a higher mean score than the SF-6D. This has to be balanced against the “floor effect” of the SF-6D, which would tend to be associated with poorer discrimination between poorer health states (e.g., lower utility states). Additionally, SF-6D appears to be more responsive.

One possible limitation of the use of IBDQ to predict utility is the possibility that the negative impact of side effects on utility may not be adequately captured. Disease-specific instruments, like the IBDQ, are designed to capture the HRQoL impacts of the disease and are not normally designed to capture the impact of treatment related side effects. Therefore, utility estimates based on the IBDQ may adequately distinguish treatments with different efficacy but be less sensitive to differences in side-effect profiles. Potential users of the algorithm should consider this possible limitation.

Based on the variance explained, the relationships between the CDAI and utilities in the simple models are weaker than those for the IBDQ and suggest that the CDAI provides a poorer basis for estimating utilities. Again its relatively poor performance as a predictor of utility reflects its main role as clinical indicator of disease activity, rather than of HRQoL.

It should be noted that our analysis uses data from patients in a multinational trial and utility values from UK population samples. This is common practice but it needs to be recognized that in some instances analysts might want to use value sets specific to other countries where such exist [18].

The ability of the preferred IBD/SF-6D algorithm to predict SF-6D values from IBDQ observations was further confirmed using an independent data set of IBDQ and SF-6D data pairs. No statistically significant difference was observed between the observed and estimated values. Given the strength, consistency, and predictable characteristics of the relationships, the algorithms appear to provide valuable and valid methods to estimate utilities from IBDQ scores in trials of Crohn’s disease patients that have collected IBDQ scores but not utilities. The generalisability of this relationship to other groups of patients, for which the IBDQ is appropriate, should be investigated.

Source of financial support: This study was funded by Elan Pharmaceuticals. The lead author retained full control over methodology and had unrestricted rights to publish the results.

References

- 1 The EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199–208.
- 2 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21:271–92.

- 3 Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single attribute utility functions for the health utilities index mark 3 system. *Med Care* 2002;40:113–28.
- 4 Sengupta N, Nichol MB, Wu J, Globe D. Mapping the SF-12 to the HUI3 and VAS in a managed care population. *Med Care* 2004;42:927–37.
- 5 Lundberg I, Johannesson M, Isacson D, et al. The relationship between health state utilities and the SF-12 in a general population. *Med Decis Making* 1999;12:128–40.
- 6 Longworth L, Buxton M, Sculpher M, Smith DH. Estimating utility data from clinical indicators for patients with stable angina. *Eur J Health Econ* 2005;6:347–53.
- 7 Best WR, Bechtel JM, Singleton JW, et al. Development of a Crohn’s disease activity index. National Cooperative Crohn’s Disease Study. *Gastroenterology* 1976;70:439–44.
- 8 Guyatt G, Mitchell A, Irvine EJ, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology* 1989;96:804–10.
- 9 Feagan BG, Yan S, Bala M, et al. The effects of infliximab maintenance therapy on health related quality of life. *Am J Gastroenterol* 2004;18:465–6.
- 10 Brazier J, Roberts J, Tsuchiya A, et al. A comparison of the EQ-5D and the SF-6D across seven patient groups. *Health Econ* 2004;13:873–84.
- 11 Konig H-H, Ulshofer A, Gregor M, et al. Validation of the Euroqol questionnaire in patients with inflammatory bowel disease. *Eur J Gastroenterol Hepatol* 2002;14:1205–15.
- 12 Sandborn WJ, Colombel JF, Enns R, et al. for the International Efficacy of Natalizumab as Active Crohn’s Therapy (ENACT-1) and the Evaluation of Natalizumab As Continuous Therapy (ENACT-2) Trial Groups. Natalizumab induction and maintenance therapy for Crohn’s disease. *N Engl J Med* 2005;353:1912–25.
- 13 Gregor JC, McDonal JW, Klar N, et al. An evaluation of utility measurement in Crohn’s disease. *Inflamm Bowel Dis* 1997;3:265–76.
- 14 Brooks R, with the EuroQol Group. EuroQol: the current state of play. *Health Pol* 1996;37:53–72.
- 15 Dolan P. Modeling valuations for EuroQol Health States. *Med Care* 1997;11:1095–108.
- 16 Ware JE, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–81.
- 17 Cheung WY, Garratt AM, Russell IT, Williams JG. The UK IBDQ—a British version of the inflammatory bowel disease questionnaire, development and validation. *J Clin Epidemiol* 2000;53:297–306.
- 18 Szende A, Oppe M, Devlin N (eds). EQ-5D value sets: Inventory, comparative review and user guide. Dordrecht, The Netherlands: Springer, 2007.