Estimating Utility Values for Health States of Type 2 Diabetic Patients Using the EQ-5D (UKPDS 62)

Philip Clarke, PhD, Alastair Gray, PhD, Rury Holman, FRCP

Purpose**. The aim of this study was to analyze quality-of-life data from the United Kingdom Prospective Diabetes Study (UKPDS) to estimate the impact of diabetes-related complications on utility-based measures of quality of life. **Methods**. The EuroQol EQ-5D instrument was administered in 1996 to 3667 UKPDS patients with type 2 diabetes. Tobit and censored least absolute deviations (CLAD) regression analysis based on data from the 3192 respondents was used to estimate the impact of major complications on (1) the visual analog scale (VAS) and (2) the EQ-5D utilities derived from population-based time trade-off values. **Results**. Using the tobit model, the effect on tariff values was as follows: myocardial infarction = –0.055 (95% confidence interval [CI] = –0.067, –0.042), blindness in 1 eye = –0.074 (95% CI = –0.124, –0.052), ischemic heart disease = –0.090 (95% CI = –0.126, –0.054), heart failure = –0.108 (95% CI = –0.169, –0.048), stroke = –0.164 (95% CI = –0.222, –0.105), and amputation = –0.280 (95% CI = –0.389, –0.170). The impact on the VAS scores was smaller, but the ranking was identical. Estimates of these effects, based on the nonparametric CLAD estimator, are also reported and compared. **Conclusion**. These results demonstrate the magnitude of the impact of 6 complications on utility-based measures of quality of life, which can be used to estimate the outcome of interventions that reduce these diabetes-related complications. **Key words:** diabetes; utility; EQ-5D; quality of life; visual analog scale; diabetes-related complications. **(Med Decis Making 2002;22:340–349)

The increasing use of quality-adjusted life years as a measure of outcome in cost-utility studies has been accompanied by a growing reliance on secondary data sources to provide estimates of the utility values that reflect preferences for various health states. Health economists have traditionally used methods such as the standard gamble and time trade-off method on patients or the general public (based on surveys involving health state descriptions) in order to collapse the quality and length of life into a single measure.[1] However, there are often limited opportunities for conducting such studies in practice, so reference values for the utility associated with particular health states are often

Received 9 February 2001 from the Health Economic Research Centre, Department of Public Health, University of Oxford, Institute of Health Sciences, Oxford, United Kingdom (PC, AG); and the Diabetes Trials Unit, Radcliffe Infirmary, Oxford, United Kingdom (RH). An earlier version of this article was presented at the U.K. Health Economists Study Group, Nottingham, July 2000. The major grants for the United Kingdom Prospective Diabetes Study were from the U.K. Medical Research Council; the British Diabetic Association; the U.K. Department of Health; the National Eye Institute; the National Institute of Digestive, Diabetes and Kidney Disease at the National Institutes of Health; the British Heart Foundation; Novo-Nordisk; Bayer; Bristol Myers Squibb; Hoechst; Lilly; Lipha; and Farmitalia Carlo Erba. We would like to thank Stirling Bryan and other participants at the Nottingham Health Economists Study Group meeting, as well as David Wright, Carole Cull, David Matthews, and Irene Stratton. We also thank the 2 anonymous referees for helpful comments on an earlier draft of this article. We received very useful statistical advice from Bill Greene and are grateful to Dean Jolliffe, whose STATA program was used to estimate the censored least absolute deviations model with bootstrapped standard errors. Any remaining errors are the authors' responsibility. This article is number 62 in the United Kingdom Prospective Diabetes Study series. Revision accepted for publication 7 February 2002.

Address correspondence and reprint requests to Dr. Clarke, Health Economic Research Centre, Department of Public Health, University of Oxford, Institute of Health Sciences, Old Road, Oxford OX3 7LF, United Kingdom; phone: 44-1865-226691; fax: 44-1865 226842; e-mail: philip.clarke@dphpc.ox.ac.uk.

employed when explicit evaluation of quality of life is required. Typical sources of information are population surveys providing normative data on usual quality-of-life levels,[2] community surveys providing quality-of-life scores in groups reporting a range of disease conditions and disabilities,[3] and disease-specific studies reporting utilities associated with, for example, different prostate cancer health states.[4]

The primary purpose of this study is to add to this literature by reporting estimates of the effect on utility values of major complications of type 2 diabetes, using data from a large trial of therapies for diabetes, the United Kingdom Prospective Diabetes Study (UKPDS).[5] We employ a regression-based approach to estimate the effect of different complications on utility after controlling for demographic variables and other complications. It has long been recognized that health care cost data often have features, such as extreme skewness, that invalidate standard approaches to statistical inference.[6] Much less attention has been paid to the statistical issues that arise when analyzing utility or quality-of-life data. Therefore, a secondary purpose of this study is to examine the merits of different statistical methods when estimating the effect of different complications on utility.

A number of previous studies have reported quality of life among patients with diabetes. These can be divided into studies showing no relation between therapies to control blood glucose and resultant quality of life,[7,8] studies showing a significant relation between therapies and quality of life,[9,10] and studies showing that the occurrence of a complication affects quality of life.[9–15] However, these studies generally use instruments or questionnaires that are not directly useful as measures of utility. The Diabetes Control and Complications Trial, for example, reported the impact of intensive versus conventional treatment on quality of life of type 1 diabetic patients as assessed by the Diabetes Quality of Life Measure, the Symptom Checklist-90R, and the SF-36.[7] Lacking a direct measure of utility, the subsequent economic evaluation from that trial confined its effectiveness measurement to life years gained.[11]

Previously reported results from the UKPDS have shown no detectable difference in quality of life between patients allocated randomly to different therapies, and have demonstrated that the recent occurrence (within 12 months) of a microvascular or macrovascular complication significantly reduced quality of life.[12] Here, we focus exclusively on the impact of complications on utility, consider the long- and short-term effects of complications, and report results for 6 prespecified clinical events.

METHODS

Clinical Trial

The UKPDS was conducted from 1977 to 1997 in 23 participating U.K. hospitals. A total of 5102 patients with newly diagnosed type 2 diabetes were recruited to the study. From this total, eligible patients were then randomized into blood glucose control and blood pressure control studies.[13,14]

Assessment of Quality of Life

Quality of life in the UKPDS was measured in 2 ways: a specially designed questionnaire examined specific quality-of-life domains in terms of cognitive mistakes, mood disturbances, symptoms, and work satisfaction,[12] and the EQ-5D instrument examined generic health-related quality of life.[15] In this study, we are concerned only with the latter. The EQ-5D is a multiattribute instrument for measuring preferences associated with an individual's health state. The instrument consists of a visual analog scale (VAS) and a descriptive system covering 5 dimensions (mobility, self-care, usual activity, pain/discomfort, anxiety, and depression), each of which has 3 levels (no problem, some problem, extreme problems).[15] Reference values for each of the 3^5 or 243 health states were estimated using a survey of the general British population.[16] The EQ-5D was administered cross-sectionally to all 3667 patients in the study in 1996 (i.e., to those who had not died or been lost to follow-up). At that point, they had been in the study for a median of 10.3 years.

Patients were given the EQ-5D and asked to complete it during routine visits to UKPDS clinics, in a quiet room with no help from nursing staff, family, or friends. Patients who did not attend clinics during the survey period were sent the questionnaire by mail, and those who did not return the questionnaire were sent up to 2 reminders. Both the 5-question descriptive health state portion and the VAS were administered; health states were subsequently allocated tariff scores based on general population time trade-off valuations of the reported health states, ranging from 1 for full health to −0.594 for severe problems in all 5 dimensions[16]; the VAS was scaled at administration from 0, indicating the worst state of health, to 100, indicating the best state of health.[17]

Clinical Events

All patients had an annual assessment to determine whether they had experienced any clinical events

within the previous year. Details of each event were recorded on data sheets, and the UKPDS center concerned was asked to provide full information on the event. The available information was then presented to the UKPDS Endpoint Adjudication Committee, where 2 clinical assessors independently classified the event into predefined categories based on the 9th revision of the International Classification of Diseases codes (see the appendix). If any disagreements could not be resolved by arbitration, the information was submitted to a panel of 3 assessors for a final decision. There were a total of 10 nonfatal clinical event categories in the UKPDS. In this analysis, we examine how quality of life is affected by 6 of these: myocardial infarction (MI), ischemic heart disease (IHD), stroke, heart failure, amputation, and blindness. The remaining 4 events were excluded from this study, either because they were defined by the occurrence of treatment for a clinical condition (e.g., retinal photocoagulation, cataract extraction), making it impossible to measure the disutility of the problem itself, or because there were insufficient numbers of events ($n < 15$) for reliable analysis (e.g., renal failure, vitreous hemorrhage).

Subjects

A total of 3302 fully or partially completed EQ-5D survey forms were returned, a response rate of 90%. Of these, 110 patients were excluded: 31 respondents did not report the date they completed the survey, 26 respondents had not completed at least 1 of the 5 EQ-5D questions, and 53 patients did not complete the VAS. The remaining 3192 patients were a mean (range) of 10.6 (5–19) years from entry to the UKPDS and had a mean age of 62.3 years at the time of the questionnaire. In line with the trial exclusion criteria,[13] patients were not admitted to the study if they had suffered a myocardial infarction in the previous year; currently had angina, heart failure, or more than 1 major vascular event; or had a concurrent illness likely to limit life. To investigate the degree to which patients in the study were representative of people with diabetes in England, we compared them with a group of 350 persons reporting diabetes as a long-standing illness in the 1996 Health Survey of England.[2] This sample from the general population was 55.7% male, had a mean age of 62.3 (SD = 14.8) years, and had a mean EQ-5D tariff value of 0.70 (SD = 0.31); the VAS was not administered in that survey.

Test-Retest Reliability

The test-retest reliability of patients' responses was assessed using a randomly selected subsample of 124 patients who repeated the EQ-5D approximately 4 months after the main survey. To control for changes in health state between the 1st and subsequent survey, patients who had experienced any of the diabetes-related complications examined in the study during the intervening period were excluded from the analysis. Kappa (κ) statistics for the 5 domains of the EQ-5D and the intraclass correlation coefficient for the tariff and VAS scores were calculated using standard methods.

Statistical Issues in the Modeling of Utility Data

Regression analysis was employed to model the relationship between tariff and VAS values and clinical events after adjusting for age, sex, and the duration of diabetes. Dolan et al.[16] noted that the time trade-off valuations that form the basis of the tariff values for the EQ-5D states are bounded by 1.0 (the score for full health).[18] In EQ-5D surveys, it is common for a significant fraction of respondents to rate themselves in full health (i.e., 11111 on the EQ-5D survey); for example, population data from the 1996 Health Survey for England showed that 52% of all respondents gave that response and were assigned the tariff value equal to 1.0. In such circumstances, it has been argued[19] that conventional linear regression analysis is inappropriate for 2 reasons. First, it implies that tariff values are continuously distributed and, hence, the probability of a tariff value exactly equal to 1.0 is very small. Second, linear regression does not restrict the tariff value to always be below 1.0. To overcome these limitations, we use a tobit model[20] with upper censoring at 1.0 in the main analysis. This assumes that

$$q_i^* = \beta' x_i + \varepsilon_i \qquad (1)$$
$$q_i = \begin{cases} q_i^* & \text{if } q_i^* < 1 \\ 1 & \text{if } q_i^* \geq 1 \end{cases},$$

where q_i^* is a latent measure of quality of life, x_i is a vector of K independent variables influencing quality of life, β is a vector of coefficients on the independent variables, ε_i is an error term that is normally distributed with constant variance denoted by $\varepsilon_i \sim N(0,\sigma^2)$, and q_i is the actual utility as measured by the tariff or VAS score. The subscript i represents an observation from the sample of N observations.

To determine the impact of various clinical events on quality of life, we analyzed the relationship between the tariff and VAS values and the 6 nonfatal clinical

events (as defined in Table 1). For simplicity, we have assumed an additive specification for the independent variables in all models. The VAS scores were numerically rescaled to generate a 0.0 to 1.0 index, and a tobit model was used for consistency. The impact of clinical events on quality of life may vary over time. For example, if the underlying disease has an acute phase, the event may only have a transient impact on quality of life. Consequently, our initial model includes 2 dummy variables for each clinical event: the 1st indicated whether the patient had experienced the event in the previous year (i.e., the year prior to the survey), and the 2nd indicated a clinical event that occurred at any time since the diagnosis of diabetes, but at least 1 year prior to the survey. The coefficient on the 1st dummy variable is therefore intended to capture any acute consequences, and the coefficient on the 2nd is intended to capture the long-term impact of each clinical event. A likelihood ratio test was used to examine whether there was a significant difference between these coefficients as a way of examining whether the effect of the event on q_i had changed between these 2 time periods. When significant differences were not found, the 2 dummy variables were combined based on the assumption that the effect of complications on utility does not vary over time.

To test the applicability of the tobit model, we employed an extensive range of specification tests, as outlined in the statistical appendix. In particular, we tested for heteroscedasticity and nonnormality in the error term, since either can result in the tobit model producing inconsistent estimates of β.[21] When specification problems arise, Powell's censored least absolute deviations (CLAD) estimator provides an alternative method of estimation.[22,23] The CLAD estimator has been shown to perform well when the distributional assumptions of the tobit model are violated.[24] The method of calculating marginal effect of various clinical events on q_i for both the tobit and CLAD models is also outlined in the appendix. All models were estimated using LIMDEP 7.0 and STATA 7.0, with a P value < 0.05 considered to be statistically significant.

Table 1. Descriptive Statistics and Definitions of Variables Used in Regression Analyses (n = 3192)

Variable	Mean (*SD*)
Quality-of-life measure	
EQ-5D social tariff	0.77 (0.27)
EQ-5D visual analog scale score	0.74 (0.19)
Characteristics	
Age (years)	62.3 (9.0)
Duration of diabetes (years)	10.6 (2.8)
Previous clinical events	
Myocardial infarction	
Average time since event (years)	4.5 (3.2)
Number of patients experiencing an event:	
In the previous year (n = 25), prior to the previous year (n = 175)	
Ischemic heart disease	
Average time since event (years)	5.4 (3.3)
Number of patients experiencing an event:	
In the previous year (n = 19), prior to the previous year (n = 171)	
Stroke	
Average time since event (years)	3.7 (3.3)
Number of patients experiencing an event:	
In the previous year (n = 13), prior to the previous year (n = 56)	
Heart failure	
Average time since event (years)	3.8 (3.2)
Number of patients experiencing an event:	
In the previous year (n = 16), prior to the previous year (n = 50)	
Amputation	
Average time since event (years)	3.8 (3.4)
Number of patients experiencing an event:	
In the previous year (n = 4), prior to the previous year (n = 15)	
Blindness in 1 eye	
Average time since event (years)	5.0 (3.8)
Number of patients experiencing an event:	
In the previous year (n = 14), prior to the previous year (n = 87)	

Note: Sixty percent of the sample was male.

RESULTS

Descriptive values and definitions of the variables used in the regression analysis are shown in Table 1. The mean tariff value based on responses to the 5 EQ-5D questions was 0.77 (*SD* = 0.27), and the mean score on the VAS was 0.74 (*SD* = 0.19). The proportion of patients who had experienced the 6 diabetes-related clinical events in the previous year or prior to the previous year, as well as the number of events, is also reported in Table 1. The most commonly diagnosed clinical event was an MI, which occurred in 6.2% of patients. In contrast, the least common was amputation, which affected 0.7% of patients. In total, 645 of these diabetes-related clinical events had been experienced by patients in this analysis, and 556 patients had experienced at least 1 such event.

Table 2. Results of Linear and Tobit Regression Analysis of the Relationship between Health State Utilities (EQ-5D tariff values and visual analog scores) and Clinical Events Occurring in the Previous Year and Prior to the Previous Year (n = 3192)

	Ordinary Least Squares Model				Tobit Model			
	EQ-5D Tariff		Visual Analog Scale		EQ-5D Tariff		Visual Analog Scale	
	β_{OLS}	*SE*	β_{OLS}	*SE*	β_{OLS}	*SE*	β_{OLS}	*SE*
Constant	0.725	(0.035)	0.674	(0.025)	0.814	(0.053)**	0.683	(0.026)**
Age	0.000	(0.001)	0.000	(0.000)	0.000	(0.001)	0.000	(0.000)
Duration of diabetes	–0.001	(0.003)	0.000	(0.001)	–0.001	(0.003)	0.000	(0.001)
Male	0.092	(0.009)**	0.061	(0.007)**	0.148	(0.014)**	0.063	(0.007)**
Previous events								
Myocardial infarction								
Previous year	–0.081	(0.052)	–0.101	(0.037)**	–0.129	(0.076)	–0.106	(0.039)**
Prior to previous year	–0.044	(0.021)**	–0.042	(0.015)**	–0.078	(0.031)**	–0.045	(0.016)**
Ischemic heart disease								
Previous year	–0.141	(0.060)**	–0.105	(0.043)**	–0.205	(0.088)**	–0.112	(0.046)**
Prior to previous year	–0.079	(0.020)**	–0.041	(0.015)**	–0.132	(0.030)**	–0.044	(0.016)**
Stroke								
Previous year	–0.131	(0.073)	–0.091	(0.052)	–0.181	(0.106)	–0.096	(0.055)
Prior to previous year	–0.199	(0.035)**	–0.069	(0.025)**	–0.269	(0.051)**	–0.073	(0.027)**
Heart Failure								
Previous year	–0.058	(0.066)	–0.003	(0.047)	–0.121	(0.096)	–0.003	(0.050)
Prior to previous year	–0.134	(0.038)**	–0.092	(0.027)**	–0.181	(0.055)**	–0.095	(0.029)**
Amputation								
Previous year	–0.451	(0.131)**	–0.109	(0.094)	–0.538	(0.188)**	–0.116	(0.100)
Prior to previous year	–0.335	(0.068)**	–0.134	(0.050)**	–0.412	(0.098)**	–0.140	(0.052)**
Blindness in 1 eye								
Previous year	–0.074	(0.070)	–0.088	(0.050)	–0.094	(0.104)	–0.093	(0.053)
Prior to previous year	–0.080	(0.029)**	–0.040	(0.020)	–0.112	(0.042)**	–0.041	(0.022)
σ					0.372	(0.006)**	0.197	(0.003)**
Pseudo-R^2 measure		0.07		0.04		0.07		0.04

Note: OLS = ordinary least squares.
*P < 0.05. **P < 0.01.

Test-Retest Reliability

Of the 124 patients involved in the test-retest reliability exercise, 6 were excluded from the analysis due to the occurrence between the surveys of 1 of the 6 clinical events that may have affected their quality of life. For the 5 domains of the EQ-5D, the κ statistics ranged from 0.59 (95% confidence interval [CI] = 0.45–0.74) for the mobility domain to 0.26 (95% CI = 0.11–0.40) for the pain domain. The intraclass correlation coefficient was 0.59 (95% CI = 0.41–0.72) for the tariff scores and 0.75 (95% CI = 0.64–0.82) for the VAS scores, and therefore fell into the categories of "good" and "excellent," respectively.[25]

Relationship between Quality of Life and Clinical Events

Table 2 reports the results of the standard linear regression and the tobit regression analyses of the relationship between the tariff values or VAS scores and experiencing 1 or more clinical events within 2 different time intervals: within the previous year (i.e., at any time within the year prior to the survey) and prior to the previous year. Given the limitations of applying linear regression to these data, we focus on the tobit models.

Concerning the impact of clinical events that occurred within the previous year, patients who had been diagnosed with IHD and amputation reported a signifi-

Table 3. Results of Tobit and Censored Least Absolute Deviations (CLAD) Regression Analysis of the Relationships between Health State Utilities (EQ-5D tariff values and visual analog scores) and Clinical Events and Patient Characteristics (n = 3192)

	Tobit Model				CLAD Model			
	EQ-5D Tariff		Visual Analog Scale		EQ-5D Tariff		Visual Analog Scale	
	β_T	*SE*	β_T	*SE*+	β_C	*SE*	β_C	*SE*+
Constant	0.814	(0.053)**	0.683	(0.026)**	0.796	(0.012)**	0.750	(0.013)**
Male	0.148	(0.015)**	0.063	(0.007)**	0.054	(0.012)**	0.050	(0.013)**
Previous events								
Myocardial infarction	–0.084	(0.029)**	–0.054	(0.014)**	–0.054	(0.022)**	–0.050	(0.017)**
Ischemic heart disease	–0.139	(0.029)**	–0.051	(0.015)**	–0.069	(0.019)**	–0.050	(0.014)**
Stroke	–0.253	(0.046)**	–0.079	(0.024)**	–0.14	(0.042)**	–0.060	(0.031)
Heart failure	–0.167	(0.048)**	–0.074	(0.025)**	–0.071	(0.022)**	–0.100	(0.036)**
Amputation	–0.4340	(0.086)**	–0.131	(0.045)**	–0.415	(0.175)**	–0.160	(0.064)**
Blindness in 1 eye	–0.114	(0.039)**	–0.050	(0.020)*	–0.071	(0.033)*	–0.050	(0.033)
σ	0.372		0.197					
Pseudo-R^2 measure		0.07		0.04		0.03		0.03
Hausman test statistic		73.66**		12.34				

*$P < 0.05$. **$P < 0.01$.

cantly lower tariff value. Prior to the previous year, macrovascular events (e.g., MI, IHD, stroke, heart failure) and amputation had a negative effect on both the tariff and VAS values whereas blindness had a significant and negative effect only on the tariff values. Gender is shown to have an influence, with men reporting health states that have higher tariff scores and higher VAS values. To test the assumption of homoscedasticity, both models were compared with more general models (where the variance was assumed to be a function of gender and the clinical events). The test statistics of 13.57 (P = 0.40) and 19.72 (P = 0.10) do not lead to the rejection of the null hypothesis of homoscedasticity. To determine whether the coefficients on each clinical event were stable over time, likelihood ratio tests were also used to test the restriction that coefficients for each of the 6 complications did not change over time. The test statistics ranged from 0.18 (P = 0.67) for blindness to 0.77 (P = 0.37) for IHD in the tariff model and from 0.04 (P = 0.84) for amputation to 2.61 (P = 0.11) for heart failure in the VAS model. Thus, events occurring more than 1 year previously did not have a significantly different impact on the tariff or VAS values than events occurring within the previous year. However, these results should be interpreted with caution because for some complications the number of patients experiencing events within the previous year is small (e.g., only 4 patients had an amputation), suggesting that there may be insufficient numbers of patients to detect clinically important differences in health status.

The 2 dummy variables representing each clinical event were combined into a single variable to indicate whether an individual had been diagnosed with the event while participating in the trial. This more parsimonious model is reported in Table 3. Again, gender had an effect, with males having significantly (at the 1% level) higher tariff and VAS values. Being diagnosed with any of the clinical events also had a significant negative impact on these scores. The Hausman test statistic H indicates that the null hypothesis (of homoscedasticity and normality) is rejected for the tariff scores but not for the VAS scores.

Effect of Clinical Events on Tariff and VAS Values

For those patients who had not experienced any of the 6 diabetes-related complications, the mean tariff was 0.785 for the tariff scores and 0.747 for the VAS scores. Table 4 reports the marginal effect of each clinical event computed at the sample mean. Given the absence of significant differences in the impact of these clinical events over time, these marginal effects are

Table 4. Marginal Effect and 95% Confidence Intervals of Clinical Events on Health State Utilities

	Tobit Model		CLAD Model
	Tariff Values	Visual Analog Scale	Tariff Values
Myocardial infarction	–0.055 (–0.067, –0.042)	–0.041 (–0.043, –0.038)	–0.035 (–0.061, –0.008)
Ischemic heart disease	–0.090 (–0.126, –0.054)	–0.044 (–0.071, –0.018)	–0.044 (–0.071, –0.018)
Stroke	–0.164 (–0.222, –0.105)	–0.069 (–0.112, –0.026)	–0.090 (–0.147, –0.032)
Heart failure	–0.108 (–0.169, –0.048)	–0.065 (–0.109, –0.021)	–0.045 (–0.082, –0.008)
Amputation	–0.280 (–0.389, –0.170)	–0.120 (–0.201, –0.038)	–0.266 (–0.476, –0.055)
Blindness in 1 eye	–0.074 (–0.252, –0.124)	–0.043 (–0.078, –0.008)	–0.045 (–0.088, –0.003)
Decomposition of marginal effect			
Fraction of sample below limit	0.64	0.90	
Fraction of mean total response due to response below the limit	0.45	0.72	

Note: CLAD = censored least absolute deviations.

based on the time invariant models reported in Table 3. Because the CLAD estimator produces significantly different estimates compared to the tobit model when applied to the tariff values (as demonstrated by the Hausman test), we report the marginal effect associated with both approaches in columns 2 and 4 of Table 4. The greatest impact on the tariff was associated with the event *amputation*, which reduced the tariff value by 0.280 based on the tobit model and 0.266 based on the CLAD model. The comparative effect on the VAS score was smaller at 0.120. It is important to note that, for the effect of events on tariff and VAS scores, multiple events are assumed to have an additive impact. For example, for a patient who has experienced an MI and has a history of IHD, using the traditional tobit model the tariff value will be reduced by 0.145, which is the sum of individual marginal effects for these conditions.

To provide insight into the nature of the change, we report 2 further statistics for the tobit models: the proportion of the sample not in full health and the fraction of the mean effect due to the response below the limit value obtained using equation (1). In the VAS model, 90% of the observations are < 1.0. For any of the clinical events, only 28% of the total change in VAS values would be generated by changes in the probability of moving away from the limit value (full health), with the remaining 72% generated by movements elsewhere along this scale. Decomposition of the marginal effect in the tariff model reveals quite a different story. In the tariff model, a much lower proportion of observations are below the limit value (64%). Decomposition shows that only 45% of the change in the tariff values are due to marginal changes in tariff values of states below full health; the majority of the change is due the reduction in the probability of being in full health.

DISCUSSION

This article reports utility values associated with major complications of type 2 diabetes. It draws on information collected from a large long-term clinical trial, such that information on the clinical history of all patients was available from the time of diagnosis of diabetes onward. To estimate the separate effect of 6 diabetes-related complications on utility, we used regression-based methods and explored several of the statistical issues that arise when analyzing these types of quality-of-life data.

A key feature of this study was that it was based on a cross-sectional survey administered toward the end of the clinical trial. This has 2 important implications for the analysis and interpretation of the results. First, the pseudo-R^2 measures for the regression models indicate a relatively low goodness-of-fit. Although this is partly because our models do not include other non-diabetes-related conditions and some confounders (e.g., income) that influence health, it may also result from intra- and interpatient variability in responses to the VAS and health state descriptors. The test-retest analysis does indicate some intrapatient variation, although the 4-month interval between surveys was longer than in many other such studies. Second, a healthy survivor effect will exist to the extent that patients experiencing complications that had a more severe effect on their quality of life are less likely to have survived to participate in this quality-of-life study. Furthermore, the higher average tariff score of patients participating in the UKPDS compared to persons in the general population reporting diabetes (as reported in the 1996 Health

Survey for England) may be due to the treatment and monitoring they received during the trial.

Applying regression models to the EQ-5D tariff and VAS scores has highlighted several important statistical issues that are also likely to arise when analyzing other quality-of-life data. The most important feature is that a considerable proportion of patients rate themselves in a state classified as "full health" and, therefore, are assigned the maximum value of 1.0. To deal with this form of censoring, we used both the traditional tobit model and the CLAD estimator, which is based on median regression. The reason for employing the latter is that although it is less efficient than the tobit model, it is robust to changes in the distribution of the error term. However, choosing between these 2 approaches should not be based on statistical considerations alone. An important issue concerns how responses from different patients should be aggregated to estimate the overall effect associated with a particular clinical event. It has been argued[1] that the theoretically correct method is to calculate the *mean* utility regardless of the degree of skewness, based on the welfare economic principle that the strength of *all* individuals' preferences should count. On these grounds, the CLAD estimator runs into difficulty because it gains the desirable statistical property of consistency through the use of median regression that does not account for the strength of all individuals' preferences. In such circumstances, we report marginal effects based on the tobit and CLAD estimators and suggest that both can be used in applied work. Furthermore, although these estimators provide a straightforward method of estimating the *marginal effect* of various clinical events on q_i, other statistical techniques such as 2-part models[26] could provide a different approach to estimation. Given the increasing availability of patient-level information from quality-of-life surveys that can be used to calculate reference utilities for health states, it would be useful to explore the relative merits of this alternative approach in future research.

A striking feature of the analysis is the strong influence on the results of changes in the proportion of patients at the maximum value of the EQ-5D, that is, patients initially indicating they are in a state of full health. As Table 4 showed, more than half of the recorded changes in mean quality-of-life scores following a clinical event can be attributed to the move away from the state "full health." The inability to record relatively small changes in quality of life using the EQ-5D instrument has been noted previously[27]: at the full-health state (11111), the minimum possible health change—from 1 to 2 in 1 dimension—corresponds to a reduction in utility of between 0.12 and 0.2. This may not be a particular problem in the present context, as all the complications considered are relatively serious in nature; however, the inability to quantify a state of health between the utility values of 0.88 and 1 may compromise the ability of the instrument to detect quality-of-life changes associated with less serious complications and may raise issues of appropriate methods of estimation.

The results reported here should be of interest to researchers interested in quality-of-life measurement, and should also be of value for future cost-utility analyses in the area of type 2 diabetes. The major published economic evaluations to date from the UKPDS have focused on life years lost and endpoint-free time as the main measures of outcome.[28,29] However, the estimation of the quality-of-life impact of complications, coupled with development of a lifetime model of risk, will allow the cost-utility of different strategies for the management of diabetes to be calculated. Other analysts should also find the results of use in estimating the cost utility of different current and future interventions aimed at reducing the complications of diabetes.

APPENDIX

Description of the Diabetes-Related Complications Included in the Regression Analysis

Myocardial infarction: diagnosed with a myocardial infarction (9th revision of the International Classification of Diseases [ICD9] Code 410)

Ischemic heart disease (IHD): diagnosed with ischemic heart disease (ICD9 Codes 411 to 414.9)

Stroke: diagnosed with a major stroke with symptoms that persist more than 1 month (ICD9 Codes 430 to 434.9 and 436)

Heart failure: diagnosed with heart failure (ICD9 Codes 428 to 428.1)

Amputation: diagnosed with major limb complications requiring amputation of digit or limb for any reason (ICD9 Codes 5.845 to 5.848)

Blindness in 1 eye: diagnosed as having blindness in 1 eye (ICD9 Codes 369 to 369.9)

Statistical Discussion

Several aspects of quality-of-life data may lead to violations of the assumptions underlying the standard tobit model. The most important of these is that the error term may be heteroscedastic and nonnormally distributed, both of which lead to tobit model estimates (denoted as β_T) that are inconsistent estimates of β.[21] Hence, it is important to test (and potentially correct) for misspecification. Heteroscedasticity occurs when the variance of the error is not constant but depends on one or more of the independent variables. One way of dealing with heteroscedasticity is to specify that $\sigma_i^2 = \sigma^2 e^{\alpha' w_i}$, where w_i is a vector of variables influencing the variance of the error term and α is a vector of coefficients for these variables. The tobit model can then be reestimated assuming $\varepsilon_i \sim N(0,\sigma_i^2)$. A likelihood ratio test can again be used to determine whether the general or the standard model (which imposes the restriction that $\alpha = 0$) should be used. A 2nd approach that has also been shown to be robust to changes in distribution of the error term is Powell's censored least absolute deviations (CLAD) estimator, which is based on median regression.[22] Although this estimator is less statistically efficient than the tobit model, it has been shown to produce consistent estimates (denoted as β_C) in the presence of heteroscedasticity and nonnormally distributed errors.[24] For these reasons, we estimate the CLAD estimator alongside the standard tobit model and apply a recently developed bootstrap procedure to estimate standard errors.[30] A useful way of testing for violations of these statistical assumptions is to compare these models using a Hausman test. [24] Under the null hypothesis of no violations, both estimators are consistent but the tobit model is more efficient, and under the alternative hypothesis, only the CLAD estimator is consistent. The test statistic is computed as[21]

$$H = (\beta_C - \beta_T)'(V_C - V_T)^{-1}(\beta_C - \beta_T),$$

where V_C and V_T are the covariance matrices of the coefficients from the CLAD and tobit estimators, respectively. This test statistic has a chi-square distribution with the degrees of freedom equal to the number of common coefficients in the models being compared.

The goodness-of-fit of the tobit model (R^2_{MZ}) was measured based on an extension of a pseudo-R^2 measure for the probit model[31]: this measure has the desirable property of being a good predictor of what R^2 would be if the data were not constrained at 1.0 and ordinary least squares was applied.[32]

The marginal effect, which we denote as $\partial E[q_i | x_i]/\partial x$, of each clinical event on the observed tariff and VAS values was then calculated using the standard formula:

$$\frac{\partial E\left[q_i | x_i\right]}{\partial x_i} = \beta\Phi\left(-\frac{\beta' x_i}{\sigma}\right).$$

The marginal effect is simply β multiplied by the probability of a nonlimit observation; using a recent result that shows this applies to any continuous distribution,[33] estimates of the marginal effect based on the CLAD estimator are derived by multiplying β_C by the observed proportion of nonlimit observations.

Given the structure of the tobit model, it is well known that the marginal effect can be decomposed into 2 parts,[34]

$$\frac{\partial E\left(q_i | x_i\right)}{\partial x_i} = \Pr\left[q_i < 1\right]\frac{\partial E\left[q_i | x_i, q_i < 1\right]}{\partial x_i} + \left[E\left[q_i, q_i < 1\right] - 1\right]\frac{\partial \Pr(q_i < 1)}{\partial x}_i,$$

where $\Pr[q_i < 1]$ is the probability of a nonlimit observation (i.e., when the individual is not in full health). An intuitive explanation of this decomposition is that the total effect of a change in x_i comprises 2 parts: the change in q_i for those below the limit weighted by the probability of not being in full health and the change in probability of not being in full health weighted by the expected value of deducted from 1.0 (i.e., the score if the patient is in full health).

REFERENCES

1. Dolan P. The measurement of health-related quality of life. In: Culyer AJ, Newhouse JP, eds. Handbook of Health Economics. Amsterdam: Elsevier, 2000. p. 1723–55.
2. Department of Health/University College London. Health Survey for England 1996. London: The Stationery Office, 1998.
3. Fryback DG, Dasbach EJ, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making. 1993;13:89–102.
4. Litwin MS, Hays RD, Fink A, et al. Quality-of-life outcomes in men treated for localized prostate cancer. JAMA. 1995;273:129–35.
5. United Kingdom Prospective Diabetes Study Group. UK Prospective Diabetes Study VIII: study design, progress and performance. Diabetologia. 1991;34:877–90.
6. Jones AM. Health econometrics. In: Culyer AJ, Newhouse JP, eds. Handbook of Health Economics. Amsterdam: Elsevier, 2000. p. 265–344.
7. Diabetes Control and Complications Trial Group. Influence of intensive diabetes treatment on quality-of-life outcomes in the diabetes control and complications trial. Diabetes Care. 1996;19:195–203.
8. Weinberger M, Kirkman MS, Samsa GP, et al. A nurse-coordinated intervention for primary care patients with non-insulin-dependent diabetes mellitus: impact on glycemic control and health-related quality of life. J Gen Intern Med. 1995;10:59–66.
9. Glasgow RE, Ruggiero L, Eakin EG, Dryfoos J, Chobanian L. Quality of life and associated characteristics in a large national sample of adults with diabetes. Diabetes Care. 1997;20:562–7.
10. Keinanen KS, Ohinmaa A, Pajunpaa H, Koivukangas P. Health related quality of life in diabetic patients measured by the Nottingham Health Profile. Diabetic Med. 1996;13:382–8.
11. DCCT. Lifetime benefits and costs of intensive therapy as practiced in the diabetes control and complications trial. JAMA. 1996;276:1409–15.
12. United Kingdom Prospective Diabetes Study Group. Quality of life in type 2 diabetic patients is affected by complications but not by intensive policies to improve blood glucose or blood pressure control (UKPDS 37). Diabetes Care. 1999;22:1125–36.
13. United Kingdom Prospective Diabetes Study Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet. 1998;352:837–53.
14. United Kingdom Prospective Diabetes Study Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). Lancet. 1998;352:854–65.
15. EuroQol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy. 1990;16:199–208.
16. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ. 1996;5:141–54.
17. Gudex C, Dolan P, Kind P, Williams A. Health state valuations from the general public using the visual analogue scale. Quality Life Res. 1996;5:521–31.
18. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35:1095–108.
19. Ruud P. An Introduction to Classical Econometric Theory. Oxford, UK: Oxford University Press, 2000.
20. Long JS. Regression Models for Categorical and Limited Dependent Variables. London: Sage, 1997.
21. Greene WH. Econometric Analysis. London: Prentice Hall, 1997.
22. Powell JL. Least absolute deviations estimation for the censored regression model. J Econometrics. 1984;25:303–25.
23. Chay KY, Powell JL. Semiparametric censored regression models. J Econ Perspect. 2001;15:29–42.
24. Deaton A. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. Baltimore, MD: Johns Hopkins University Press, 1997.
25. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86:420–8.
26. Mullahy J. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. J Health Econ. 1998;17:247–81.
27. Jenkinson C, Gray A, Doll H, Lawrence K, Keoghane S, Layte R. Evaluation of index and profile measures of health status in a randomized controlled trial: comparison of the Medical Outcomes Study 36-Item Short Form Health Survey, EuroQol, and disease specific measures. Med Care. 1997;35:1109–18.
28. United Kingdom Prospective Diabetes Study Group. Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetes: UKPDS 40. Br Med J. 1998;317:720–6.
29. United Kingdom Prospective Diabetes Study Group. Cost effectiveness of an intensive blood glucose control policy in patients with type 2 diabetes: economic analysis alongside a randomised controlled trial: UKPDS 41. Br Med J. 2000;320:1373–8.
30. Jolliffe D. Skills, schooling, and household income in Ghana. World Bank Econ Rev. 1998;12:81–104.
31. McKelvey R, Zavoina W. A statistical model for the analysis of ordinal level dependent variables. J Mathematical Sociol. 1975;4:103–20.
32. Veall MR, Zimmermann KF. Goodness of fit measures in the tobit model. Oxford Bull Econ Stat. 1901;56:485–99.
33. Greene W. Marginal effects in the censored regression model. Econ Lett. 1999;64:43–9.
34. McDonald JF, Moffitt RA. The uses of tobit analysis. Rev Econ Stat. 1980;62:318–21.