Estimation of Patient Preference-Based Utility Weights from the Functional Assessment of Cancer Therapy—General

Deborah Dobrez, PhD,[1] David Cella, PhD,[2] A. Simon Pickard, PhD,[3] Jin-Shei Lai, PhD,[2] Angel Nickolov, MS[4]

[1]School of Public Health, University of Illinois at Chicago, Chicago, IL, USA; [2]Center on Outcomes, Research, and Education, Evanston North-western Healthcare, Evanston, IL, USA; [3]Center for Pharmacoeconomic Research and Department of Pharmacy Practice, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA; [4]Mallinckrodt/Tyco Healthcare, St. Louis, MO, USA

ABSTRACT

Objective: The goal of this study was to estimate an algorithm to convert responses to the Functional Assessment of Cancer Therapy—General (FACT-G) to time trade-off (TTO) utilities based on utilities for current health elicited from cancer patients.
Methods: Data for 1433 cancer patients were randomly separated into construction and validation samples. Four FACT-G questions were selected for inclusion based on correlation with Eastern Clinical Oncology Group—Performance Status (ECOG-PS) scores and TTO utilities. Item response theory was used to collapse response categories. Ordinary least squares regression with the constant constrained to one was used to estimate the algorithm.
Results: The algorithm estimated mean utility for the full validation sample within three points of observed mean utility (0.805 vs. 0.832, $P < 0.01$). Mean utilities were well predicted (mean absolute difference < 0.03, P-values > 0.05) for most subgroups defined by ECOG-PS and Short Form-36 physical functioning scores, and responses to the FACT-G overall quality of life item. Nevertheless, the algorithm systematically overpredicted utilities for poorer health states.
Conclusions: A FACT-G-based algorithm of cancer patient utilities was developed that estimates group mean utility scores with accuracy comparable to other indirect preference-based measures of health-related quality of life. Patient-based preferences for health outcomes of cancer treatment may be useful in multiple situations, such as managing resources within cancer centers and in understanding health states preferences among cancer experienced patients before and after treatment.
Keywords: cancer, quality of life, time trade-off, utility.

Introduction

The EQ-5D [1,2] and Health Utilities Index [3,4] are generic measures of health-related quality of life (HRQL) with societal reference weights for their classification systems that can help to inform health-care resource allocation. Some measures, like the Short Form (SF)-36 and SF-12, have been adapted for this purpose as well [5–7]. In addition, a set of societal preference-weights were recently derived for the Functional Assessment of Cancer Therapy—Lung Cancer [8], an HRQL measure often included in clinical trials with a quality of life end point [9].

The use of societal preference weights in HRQL measurement is supported by the notion that it is the general population, that is, (tax) payer, who ultimately pays for a given medical technology, so their preferences should be used in reimbursement decisions. This argument is most tenable in a socialized health-care environment. In privatized health-care systems, the ability to calculate utility-based HRQL scores derived from cancer-experienced individuals may be an important complement to societal-based preference weights in informing decision-making, particularly for clinical decisions in cancer. Because patient preferences are known to differ systematically from community preferences, information regarding patient preferences can help the clinician describe how patients feel about a treatment option, and determine what the optimal treatment decision would be for an individual or group of patients at his/her clinic. Finally, the use of patient preferences can help in identifying when a patient is an outlier relative to group-based preferences.

In the field of oncology, the EORTC-QLQ30 and the Functional Assessment of Cancer Therapy—General (FACT-G) are the most widely translated and used cancer-focused HRQL instruments. These measures are frequently incorporated in clinical trials of cancer therapies. Nevertheless, neither instrument provides a preference-based score. The purpose of our study was to understand which aspects of the FACT-G [10] were significant predictors of cancer-experienced patient utilities. The FACT-G

1

Address correspondence to: Deborah Dobrez, Health Policy and Administration (M/C 923), School of Public Health, University of Illinois at Chicago, 1603 W. Taylor Street, Chicago, IL 60612, USA. E-mail: ddobrez@uic.edu

consists of items and domains that are considered important to cancer-experienced patients. The results of this study delineate a subset of FACT-G items and levels related to utility (based on the time trade-off [TTO]) in terms of willingness to trade life-years as expressed by patients. The resultant algorithm may be useful as an alternative or complement to societal preference-weights when comparing outcomes of treatment in cancer either for clinical or economic decision-making purposes.

Methods

This study is a retrospective study of data collected prospectively for a prior quality of life study (National Cancer Institute R01 CA60068). The algorithm was based on directly elicited TTO utilities provided by a large sample of cancer patients for their current health state and who also completed the FACT-G. The construction and testing of the algorithm to map FACT-G responses onto TTO utilities was conducted in four steps. First, the eligible sample was randomly divided into algorithm construction and validation samples of equal size. Using the construction sample, FACT-G questions and response categories were then selected to maximize the model's expected predictive ability over a wide range of utility scores. Multiple regression models were explored to test for possible differences in predictive ability. The selected model was estimated using the construction sample. Finally, out-of-sample predictive ability was estimated using multiple groupings of subjects in the validation sample.

Data were drawn from the previously collected sample of 1714 cancer and HIV/AIDS patients who participated at five academic medical centers: Rush-Presbyterian-St. Luke's Medical Center, Robert H. Lurie Cancer Center of North-western University, Fox Chase Cancer Center, Johns Hopkins Oncology Center, and Medical College of Ohio [11]. Inclusion criteria for the original prospective quality of life study were broad, and included able to read and speak English, and diagnosis of cancer or HIV. Demographic and clinical data were collected, in addition to five health status questionnaires (two of which were used in this article), and a TTO instrument. Each participant completed questionnaires only once. To maximize the generalizability of the model, the additional exclusion criteria applied to the sample for this retrospective analysis were few, and included: 1) failure to complete either the TTO and/or FACT-G questionnaires; 2) failure to comprehend the TTO, as judged by the interviewer; and 3) diagnosis of HIV. A small number of subjects (21 in the full sample) stated that their current health was worse than death, suggesting a utility < 0. Utilities less than zero were set equal to zero for this analysis.

Individuals were offered a choice between 1 year in their current health defined as their health over the previous 2 weeks, or a specified amount of time less than 1 year in perfect health. The amount of time in perfect health was incrementally lowered until respondents were indifferent between the year spent in their current health state, and the specified length of time in perfect health. The TTO was interviewer administered, and a visual aid was used to assist in subject understanding.

The FACT-G has established reliability and validity, including sensitivity to change [10]. Summary scores may be calculated for four dimensions: physical social, emotional, and functional well-being, in addition to the overall total score. Each question has five possible responses and can be scored so that a '4' indicates no problems and a '0' indicates worst possible problems with respect to the particular item.

In addition to the TTO and FACT-G, patient-rated Eastern Clinical Oncology Group—Performance Status (ECOG-PS) was used in the selection of questions, and both ECOG-PS and selected SF-36 questions were used in the model validation. These two instruments were selected because they are primary patient function and health status questionnaires typically used in oncology clinical trials. The implication for quality of life and current health utility of additional variables, such as stage of disease and treatment status, vary considerably across diagnoses, and were therefore not used in the validation.

Questions were selected for inclusion in the algorithm based in part on their correlation with the TTO and ECOG-PS scores. Questions were also selected to represent the full range of observed quality of life, by ranking questions according to their mean scores, with lower mean scores (representing worse health) receiving some preference for inclusion. Most questions did not consistently perform best or worse by all criteria; questions were therefore selected based on a subjective application of the criteria. For example, a lower correlation with the TTO was considered acceptable to ensure that a question with a low mean score was included, to better predict utilities for poor health. One question (SWB—"am sexually active") was excluded from consideration because of an observed high number of missing responses in studies (subjects are given the option to skip the question). In addition, the final question in the FACT-G (functional well-being [FWB]—"am content with my QOL right now" was excluded, because it is a global measure that is strongly correlated with the remaining FACT-G questions. Correlations between FACT-G questions were used to identify questions that appeared to independently predict utility.

Each FACT-G question has five possible response categories, ranging from "not at all" to "very much." Response categories were combined to reduce the

number of variables in the model. We examined the rating scale structure of each item using a logistic regression-based item response theory model and WIN-STEPS [12], a computer software program for Rasch measurement models. An acceptable rating scale structure was determined based on a number of criteria: fit statistic (mean square ≤ 1.4), ordered step calibration, nonskewed frequency distribution, and average measure of categories. Misfitting response categories were combined with other response categories where appropriate. Response categories were further combined as needed to ensure that the scales are monotonic with respect to utility in the regression model.

Because the goal of the algorithm was to predict mean utility, an ordinary least squares model was conducted using the construction sample data to estimate the mapping algorithm. Responses to the selected FACT-G questions made up the full set of independent variables. Each response was entered in to the model as a dummy variable, with the best possible response set as the reference case. Each variable coefficient therefore represents a decrement in quality of life relative to the best possible response on each item. To avoid a disease labeling effect that could result from an algorithm that predicts utility less than perfect health even when there are no problems reported on the FACT questions, the constant was constrained to one. Alternative model specifications were also conducted, including the use of a quasi maximum likelihood logistic model [13], and the use of mean utilities collected for groups of patients with identical FACT responses to the included questions.

The predictive ability of the algorithm was tested in multiple ways using the portion of the data set aside for validation. First, out-of-sample mean absolute errors were calculated for each model. Next, subjects in the validation sample were grouped according to a number of variables that might be used to describe patient outcomes in clinical trials or observational studies, including SF-36 physical functioning and mental health scales (with groups selected to ensure a minimum of 30 subjects per group), and responses to the ECOG-PS and general quality of life item of the FACT-G. Statistical tests were conducted to identify mean errors that were statistically different from zero.

Minimum clinically important differences (MCIDs) have been either implicitly or explicitly proposed for several generic indirect utility-based measures, with fairly consistent proposed MCID benchmarks around 0.03 on a scale where 0 is dead and 1 is the upper end point of the scale [14]. A difference of 0.03 has been considered important on the Health Utilities Index Mark 2 and Mark 3 [15] and a change or difference score of 0.033 or more has been proposed as meaningful for the SF-6D [16]. Any change in the level reported on the EQ-5D been cited as potentially meaningful [14], with the smallest coefficient being approximately 0.03 in the widely applied Dolan algorithm [2]. Given the precedents in the literature, we considered a difference of 0.03 to be meaningful.

Results

A total of 1714 subjects participated in the overall project, of which 170 subjects did not have a diagnosis of cancer. An additional 111 subjects did not complete the interview, or were judged by the interviewer to be unable to comprehend the interview. The remaining sample included 1433 subjects, with one of 10 different diagnoses, with other known cancer (n = 288), breast cancer (n = 250), prostate cancer (n = 189), colon cancer (n = 170), nonsmall cell lung cancer (n = 146), head and neck cancer (n = 164), non-Hodkins lymphoma (n = 148), Hodkins (n = 38), small cell lung cancer (n = 35), and unknown primary (n = 12). The sample was 53% male, with an average age of 57 years (range 17–99 years). The sample was racially diverse, including 83% white, non-Hispanic subjects, 13% African American, non-Hispanic subjects, and 3% Hispanic subjects. The sample was randomly stratified in to two groups, with responses from 717 subjects used to construct the mapping algorithm, and responses from the remaining 716 subjects used for validation of the algorithm.

Table 1 displays the selected FACT-G questions, along with their characteristics and rating scale structure.

After a series of analyses, two items were revised to a two-point rating scale, one to a three-point rating scale, and one to a four-point rating scale. Overall, the

Table 1 Selected question characteristics

FACT-G item	Mean score		Correlation with ECOG-PS		Correlation with TTO		Response category
	Value	Rank	Value	Rank	Rank	Value	From 01234 to
PWB: lack of energy	2.29	1	−0.52	2	1	0.27	00112
PWB: feel sick	3.24	20	−0.44	1	6	0.29	01111
FWB: able to work	2.40	3	−0.47	10	4	0.22	00111
FWB: able to enjoy life	2.94	9	−0.36	6	10	0.24	00123

ECOG-PS, Eastern Clinical Oncology Group—Performance Status, as reported by patients; FACT-G, Functional Assessment of Cancer Therapy—General; FWB, functional well-being; PWB, physical well-being; TTO, time trade-off utility.

Table 2 Regression results

FACT-G item	Coefficient	Standard error	95% confidence interval
PWB: lack of energy			
0–1 vs. 4	−0.22*	0.03	−0.28, −0.16
2–3 vs. 4	−0.11*	0.02	−0.15, −0.07
PWB: feel sick			
0 vs. 1–4	−0.15†	0.06	−0.28, −0.03
FWB: able to work			
0–1 vs. 2–4	−0.04‡	0.03	−0.09, 0.01
FWB: able to enjoy life			
0–1 vs. 4	−0.13†	0.04	−0.20, −0.05
2 vs. 4	−0.06‡	0.03	−0.13, 0.00
3 vs. 4	−0.03	0.03	−0.09, 0.02
R^2	0.17		
MAE	0.19		

*$P < 0.01$, †$P < 0.05$, ‡$P < 0.10$.
FACT-G, Functional Assessment of Cancer Therapy—General; FWB, functional well-being; MAE, mean absolute error; PWB, physical well-being.

number of possible response categories was reduced from 30 to 11, describing a total of 48 possible health states.

Table 2 describes the regression results. All coefficients were negative, with greater reductions in FACT-G responses (e.g., move from 4 to 0 vs. 4 to 3) associated with greater reductions in utility. Most decrements in quality of life described by the FACT-G

items were associated with statistically significant reductions in utility and each FACT-G response changes was associated with a clinically important reduction in utility of three percentage points or more.

The predictive ability of the model was tested by estimating utilities for the 717 subjects that were set aside. Validation sample predicted utilities are shown in Table 3. Observed and predicted mean utilities are within the three percentage points defined as the MICD (0.805 and 0.832, respectively), but the difference is statistically significant ($P < 0.01$). The range and standard deviation for the predicted utilities is substantially less than for observed utilities (0 to 1, standard deviation = 0.284 and 0.456 to 1, standard deviation = 0.121, respectively), reflecting over prediction of utilities for poor health. Observed and predicted utilities are reasonably well correlated ($r = 0.32$, $P < 0.01$).

The predictive ability of the models was further examined by calculating mean predicted utilities for subjects grouped by various criteria. Subjects were grouped according to ECOG-PS, by responses to the FACT-G question, "I am content with the quality of my life right now," and two of the eight scales defined by the SF-36 (physical functioning and mental health). For the SF-36 scales, subjects were grouped for each

Table 3 Mean prediction error by group

Groups	Mean prediction error	Groups	Mean prediction error
ECOG		FACT-G: I am content with the quality of my life right now	
Confined to bed (n = 7)	0.342*	Not at all (n = 90)	0.105*
Bed > 50% (n = 79)	0.104*	A little bit (n = 69)	−0.025
Bed < 50% (n = 233)	0.042*	Somewhat (n = 170)	0.012
Some symptoms (n = 189)	0.007	Quite a bit (n = 172)	0.016
No symptoms (n = 209)	−0.010	Very much (n = 184)	−0.004
SF-36: physical functioning (PF)		SF-36: mental health (MH)	
$0 \leq PF < 10$ (n = 30)	0.139*	$0 \leq MH < 40$ (n = 32)	0.004
$10 \leq PF < 20$ (n = 30)	0.107†	$40 \leq MH < 50$ (n = 54)	−0.006
$20 \leq PF < 30$ (n = 42)	0.006	$50 \leq MH < 60$ (n = 53)	0.024
$30 \leq PF < 40$ (n = 56)	0.011	$60 \leq MH < 70$ (n = 138)	0.026
$40 \leq PF < 50$ (n = 52)	0.042	$70 \leq MH < 80$ (n = 92)	0.010
$5 \leq PF < 60$ (n = 85)	0.027	$80 \leq MH < 90$ (n = 188)	0.011
$60 PF < 70$ (n = 60)	0.019	$90 \leq MH < 100$ (n = 105)	0.017
$70 \leq PF < 80$ (n = 87)	0.013		
$80 \leq PF < 90$ (n = 94)	−0.004		
$90 \leq PF < 100$ (n = 95)	0.000		

*$P < 0.05$; †$P < 0.10$.
ECOG, Eastern Clinical Oncology Group; FACT-G, Functional Assessment of Cancer Therapy—General; SF-36, Short Form-36.

according to their scale score in 10-point increments, with groups combined as appropriate to ensure 30 subjects per group.) Prediction results are shown in Table 3.

Mean utilities are well predicted (difference less than three percentage points, and difference not statistically significant) for most subject subgroups. Mean utilities for subject groups defined by the relatively broad ECOG-PS measure differs significantly for the three lowest functioning subgroups. Differences for the remaining three measures are not statistically and clinically meaningful for any of the subgroups except for the two lowest functioning by the SF-36 physical functioning measure and for the lowest functioning by the FACT-G overall quality of life item. In most cases, the algorithm overpredicts utilities. This is especially true for the poor health subgroups.

Use of the algorithm to estimate utilities requires that response categories first be collapsed, and converted in to sets of dummy variables for application of the regression equation. This is accomplished using the cross-walk equation below. All variables must be ordered so that a 0 indicates the worst possible response, so that two of the selected questions (physical well-being [PWB]: lack of energy and PWB: feel sick) should be reversed first.

$$Utility = 1 + \begin{pmatrix} -0.2222 & \text{if } q1 = [0, 1] \\ -0.1137 & \text{if } q1 = [2, 3] \end{pmatrix} +$$
$$(-0.1537 \quad \text{if } q2 = 0) +$$
$$(0.0431 \quad \text{if } q3 = [0, 1]) +$$
$$\begin{pmatrix} -0.1254 & \text{if } q4 = [0, 1] \\ -0.0641 & \text{if } q4 = 2 \\ -0.0345 & \text{if } q4 = 3 \end{pmatrix}$$

Where q1 = PWB: lack of energy, q2 = PWB: feel sick, q3 = FWB: able to work, and q4 = FWB: able to enjoy life. Group mean utilities can be estimated by first calculating individual predicted utilities, and then averaging within the group. Because this is a linear model, group means can also be calculated by multiplying each coefficient by its percentage frequency (i.e., if responses are evenly divided by three for the first question, each coefficient would be multiplied by one-third). STATA code to estimate utilities is available by request.

Conclusions

The derivation of a set of utility-based weights based on the FACT-G descriptive system serves multiple purposes. First, an algorithm that generates a utility-based single summary score from responses to the FACT-G enables cost-utility analysis to be performed on both retrospective and prospective FACT-G data sets. Second, prospective data collection may be conducted

without the inclusion of direct preference elicitation methods, such as the standard gamble or TTO, which some subjects find difficult to understand [17–19].

It is important to note that the US Panel suggested the use of community preferences for health states as "the most appropriate ones for use in a Reference Case analysis [20]." Societal preference weights are desirable for decisions of allocative efficiency across healthcare systems. Nevertheless, when allocating resources across cancer-based settings or as a complement to societal weights, it may be desirable to have a set of patient-based weights.

In this article, we have described the construction and validation of an algorithm to estimate cancer-experienced patient preferences for health states defined by observed responses to the FACT-G. The algorithm performed well in predicting mean utility in a set-aside validation sample by multiple measures, especially in the mid to high ranges of health. Mean predicted values were generally quite close to observed mean utilities for subject groupings defined by responses to four HRQL items.

The use of patient preferences, which are known to be higher than community preferences, has the disadvantage empirically of reducing the variation of observed utility, and limiting the number of FACT-G items and response levels that can be included with statistically and clinically meaningful coefficients. Nevertheless, the systematic overestimation of utility for poor health states demonstrated in this study has also been reported previously for algorithms that were based on community preferences [7–12].

In this study, actual observed utilities for current health were used rather than utilities for hypothetical health states derived from the FACT-G questions. This has the advantage of enabling us to predict actual quality of life as measured by directly elicited TTO utilities for current health, which is a goal of many studies that cannot otherwise include direct elicitation of utilities in the battery of assessments. A disadvantage is that it results in greater individual prediction error (as evidenced by the high mean absolute error). Because group means are typically used in cost-utility analyses and treatment comparisons, individual prediction error is of less concern than any systematic error predicting group means. Group mean prediction was very accurate, usually within a few percentage points of the directly observed TTO utility. Model group mean error overall and for multiple subgroups were usually within the range of differences commonly observed between directly elicited utilities [21–23]. 3 Therefore, mapping FACT-G scores to utilities for calculation of group mean utilities seems justified by the current study.

An additional strength of this study was that we had a large enough sample size to partition the data to include a validation sample. This allowed us to test the

out-of-sample predictive ability of our models overall, and for previously defined subsets of patients, demonstrating mean errors that were within the range of those found in the Lawrence and Fleishman (2004) models, which also tested out-of-sample predictive ability for groups of patients overall, and by selected characteristics [24].

Ultimately, the selection of FACT-G questions was determined based on multiple factors; it is possible that others would make different decisions using the same data, potentially impacting the predictive ability of the resulting model. Efforts to replicate the question selection in separate samples would be useful. Although the model was estimated using cancer patients at all stages of disease and treatment, the number of people with self-reported very poor health was small, with only 12 subjects (<1%) reporting an ECOG-PS of 4 (confined to bed) in the overall analytic sample. Oversampling of patients with relatively poor health would be helpful in the future to improve the predictive ability of the model for low utilities. Finally, the original assessment of utility did not allow for values of current health that were worse than death. Responses for subjects who stated this preference were arbitrarily assigned a score of −1, with no attempt made to elicit strength of preference. For this study, their scores were set equal to 0. Nevertheless, there is considerable evidence that some states of health are viewed as worse than death [25,26], and some preference-based measures of utility, such as the EQ-5D, explicitly allow for this in their estimation of utility. Rescaling of utility scores to allow for valuations of states worse than death would alter the mapping algorithm, but could not be explored with the data available for this study.

In conclusion, although the use of individual utilities predicted from FACT-G responses cannot be supported by these results, estimates of group mean utility scores based on FACT-G responses in a cohort of patients is justified by its accuracy as compared with applications with other instruments. The algorithm draws on data collected from cancer and so is appropriate for the estimation of current health utilities for patients. The algorithm greatly expands the pool of data available for use in cost-utility analyses and treatment comparisons: FACT-G responses from previously conducted observational studies and clinical trials can be used to estimate utilities; and response burden can be minimized in newly conducted studies without the exclusion of utilities, simply by including the FACT-G in the assessment protocol. Future research that focuses on the estimation of a societal set of preference weights for the FACT-G will help to complement the FACT-G algorithm based on cancer patient preferences presented in this article.

References

1 Kind P. The EuroQoL instrument: an index of health related quality of life. In: Spilker B, ed. Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia, PA: Lippincott-Raven, 1996.
2 Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
3 Feeny D, Torrance GW, Furlong W. Health utilities index. In: Spilker B, ed. Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia, PA: Lippincott-Raven Publishers, 1996.
4 Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. Ann Med 2001;33:375–84.
5 Lundberg L, Johannesson M, Isacson DG, Borgquist L. The relationship between health-state utilities and the SF-12 in a general population. Med Decis Making 1999;19:128–40.
6 Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol 1998;51:1115–28.
7 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.
8 Kind P, Macran S. Eliciting social preference weights for functional assessment of cancer therapy-lung health states. Pharmacoeconomics 2005;23:1143–53.
9 Cella DF, Bonomi AE, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Lung (FACT-L) quality of life instrument. Lung Cancer 1995;12:199–220.
10 Cella DF, Tulsky DS, Gray G, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993;11:570–9.
11 Cella D, Chang CH, Wright BD, et al. Defining higher order dimensions of self-reported health: further evidence for a two-dimensional structure. Eval Health Prof 2005;28:122–41.
12 Linacre JM. WINSTEPS: Rasch Analysis for All Two-Facet Models. Chicago, IL: MESA Press, 2003.
13 Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J App Econ 1996;11:619–32.
14 Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. Qual Life Res 2005;14:207–19.
15 Drummond M. Introducing economic and quality of life measurements into clinical studies. Ann Med 2001;33:344–9.
16 Walters SJ, Brazier JE. What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. Health Qual Life Outcomes 2003;1:4.
17 Hollingworth W, Deyo RA, Sullivan SD, et al. The practicality and validity of directly elicited and SF-36

derived health state preferences in patients with low back pain. Health Econ 2002;11:71–85.

18 Woloshin S, Schwartz LM, Moncur M, et al. Assessing values for health: numeracy matters. Med Decis Making 2001;21:382–90.

19 Dobrez DG, Calhoun EA. Testing subject comprehension of utility questionnaires. Qual Life Res 2004;13:369–76.

20 Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.

21 Krabbe PF, Essink-Bot ML, Bonsel GJ. The comparability and reliability of five health-state valuation methods. Soc Sci Med 1997;45:1641–52.

22 Smith DS, Krygiel J, Nease RF Jr, et al. Patient preferences for outcomes associated with surgical management of prostate cancer. J Urol 2002; 167:2117–22.

23 Jansen SJ, Kievit J, Nooij MA, Stiggelbout AM. Stability of patients' preferences for chemotherapy. the impact of experience. Med Decis Making 2001;21:295–306.

24 Lawrence WF, Fleishman JA. Predicting EuroQoL EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. Med Decis Making 2004;24:160–9.

25 Macran S, Kind P. "Death" and the valuation of health-related quality of life. Med Care 2001;39:217–27.

26 Patrick DL, Starks HE, Cain KC, et al. Measuring preferences for health states worse than death. Med Decis Making 1994;14:9–18.

AUTHOR QUERY FORM

Dear Author,

During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.

Many thanks for your assistance.

Query References	Query	Remark
q1	Au: Please note that 'HRQOL' has been changed to HRQL for spelling consistency.	
q2	Au: Please spell out SWB.	
q3	Au: Please note that the references have been reordered from number **21** onwards.	
q4	Au: **Lawrence et al. (2004)** has been changed to **Lawrence and Fleishman (2004)** so that this citation matches the list; please check.	