Volume 12 • Number 6 • 2009
VALUE IN HEALTH

Preliminary Validation of an Optimally Weighted Patient-Based Utility Index by Application to Randomized Trials in Breast Cancer

Peter S. Grimison, BSc, MBBS, MPH FRACP,[1] R. John Simes, BSc, MBBS, MD SM FRACP,[1] H. Malcolm Hudson, BSc, PhD,[1,2] Martin R. Stockler, MBBS, MSc FRACP[1,3]

[1]NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia; [2]Department of Statistics, Macquarie University, Sydney, Australia; [3]Sydney Cancer Centre, Royal Prince Alfred Hospital, Camperdown, Australia

ABSTRACT

Objectives: To optimize, apply, and validate a scoring algorithm that provides a utility index from a cancer-specific quality of life questionnaire called the Utility-Based Questionnaire-Cancer (UBQ-C) using data sets from randomized trials in breast cancer. The index is designed to reflect the perspective of cancer patients in a specific clinical context so as to best inform clinical decisions.

Methods: We applied the UBQ-C scoring algorithm to trials of chemotherapy for advanced (n = 325) and early (n = 126) breast cancer. The algorithm converts UBQ-C subscales into a subset index, and combines it with a global health status item into an overall HRQL index, which is then converted to a utility index using a power transformation. The optimal subscale weights were determined by their correlations with the global scale in the relevant data set. The validity of the utility index was tested against other patient characteristics.

Results: Optimal weights (range 0–1) for the subset index in advanced (early) breast cancer were: physical function 0.20 (0.09); social/usual activities 0.23 (0.25); self-care 0.04 (0.01); and distresses 0.53 (0.64). Weights for the overall HRQL index were health status 0.66 (0.63) and subset index 0.34 (0.37). The utility index discriminated between breast cancer that was advanced rather than early (means 0.88 vs. 0.94, $P < 0.0001$) and was responsive to the toxic effects of chemotherapy in early breast cancer (mean change 0.07, $P < 0.0001$).

Conclusions: The scoring algorithm for the UBQ-C utility index can be optimized in different clinical contexts to reflect the relative importance of different aspects of quality of life to the patients in a trial. It can be used to generate sensitive and responsive utility scores, and quality-adjusted life-years that can be used within a trial to compare the net benefit of treatments and inform clinical decision-making.

Keywords: cancer, health-state utility, health-related quality of life, patient-derived preferences.

Introduction

The quality-adjusted life-year (QALY) approach is a useful way to compare cancer treatments, because it integrates the beneficial and harmful effects of treatment on health-related quality of life (HRQL), expressed as a utility, with the effects of treatment on survival [1–3]. Analyses of cancer trials in terms of utilities and QALYs are increasingly used to inform economic decisions about cancer treatments [4–10], but can also be used to inform clinical decisions [11–16].

A practical and feasible approach to obtain utility scores for generating QALYs in cancer trials is to use a utility-based instrument. A utility-based instrument uses a scoring algorithm to convert the responses from a questionnaire that elicits ratings about various dimensions of HRQL to a utility index [1,17,18]. The scoring algorithm is valued in a valuation survey, where a sample of people directly assign a utility score to the health states described by the questionnaire using a time trade-off interview or related technique [19]. A utility-based instrument may include generic or disease-specific questions, and the scoring algorithm may generate utilities that are based on the perspective of lay people or patients. Three of the most commonly used instruments are the EuroQol EQ-5D [20], Health Utilities Index (HUI3) [21], and the Short Form 6D (SF-6D) [22]. These instruments include generic questions applicable to any disease or population, and their scoring algorithms are based on the perspective of lay people. Utility-based instruments reported more recently have included disease-specific questions and use scoring algorithms that are based on the perspective of patients rather than lay people [23–25].

Ideally, the perspective from which a utility instrument is valued should reflect the views of the population that the researcher is trying to reflect in the decision-making [1,17,26,27]. In a companion paper we emphasized that patients typically assign a higher utility to a health state than a lay person, which can have significant implications for health funding, policy, and clinical decisions that incorporate utilities and QALYs [23]. Researchers using utilities to inform health funding and policy decisions will generally prefer the perspective of lay people [28–30], whereas researchers using utilities to inform clinical decisions will generally prefer the perspective of patients [1,17,26,31,32]. This is because the objective of clinical decisions is to maximize health for an individual patient with that disease [23]. Recently, it has been recognized that the preferences and attitudes of lay people in different countries may differ, because of differences in demographic background, social and cultural values, and political and economic systems [33,34]. As a result, some scoring algorithms for utility-based instruments based on the perspective of lay people have been optimized for use in different countries to reflect these differences [33,35–37]. It has also been recognized that the preferences and attitudes of cancer patients in different clinical contexts may differ, because patients with different cancer diagnoses, stages of disease, and treatment

Address correspondence to: Peter Grimison, NHMRC Clinical Trials Centre, University of Sydney, Locked Bag 77, Camperdown NSW 2050 Australia. E-mail: Peter.Grimison@ctc.usyd.edu.au

10.1111/j.1524-4733.2009.00536.x

may assign different importance to different aspects of HRQL [38–41]. We posit that scoring algorithms for utility-based instruments based on the perspective of patients should be optimized for different treatment contexts to reflect these differences.

Lumley et al. have developed a novel approach to optimizing scoring algorithms for different clinical contexts using the HRQL data collected in that context [38]. Lumley's approach requires a questionnaire including items about the specific aspects of HRQL and a single-item global scale. We define a single-item global scale as one asking respondents directly for a unified assessment of their HRQL. Lumley's approach gives extra weight to the responses about specific aspects of HRQL that are more highly correlated with the responses on the global scale. These weights are intended to reflect the relative importance that the subjects assign to different aspects of HRQL. The optimization of the scoring algorithm requires weighting to be determined for each clinical context but does not require the valuation survey to be repeated.

The aim of this work was to use Lumley's approach to derive an optimized scoring algorithm for a cancer-specific HRQL instrument that is based on the perspective of cancer patients. In a companion paper we described the development and preliminary validation of the algorithm [23]. This paper describes the application of the algorithm to trial data sets, and illustrates how it can be optimized in different treatment contexts.

Methods

Sources of Data

The data used to optimize, apply, and validate the scoring algorithm were collected in two randomized clinical trials of chemotherapy for breast cancer. Both studies were approved by the human research ethics committees at all participating institutions. All patients provided written informed consent.

The first trial, referred to as the "advanced cancer trial," was conducted by the Australian New Zealand Breast Cancer Trials Group. It included patients with advanced breast cancer who were randomly allocated to receive either daily oral capecitabine or standard CMF as first-line chemotherapy until disease progression [42]. The primary outcome measure of the trial was quality-adjusted time to progression. Secondary outcome measures were time to progression, response rates, HRQL, overall survival, safety, and cost-effectiveness. Eligible subjects were 18 years or older, and were about to start first-line chemotherapy for histologically confirmed advanced breast cancer. Subjects were excluded if they were totally confined to bed and completely disabled [Eastern Cooperative Oncology Group (ECOG) performance status 4, as described in the next section]. Enrolment was from June 2001 to July 2005 at 34 centers in Australia and New Zealand. Subjects completed the Utility-Based Questionnaire-Cancer (UBQ-C) and other questionnaires about HRQL that are described in the following discussion (unless they could not read English). The data described in this paper came from baseline questionnaires competed before randomization.

The second trial, referred to as the "early cancer trial," was conducted by the Australian New Zealand Breast Cancer Trials Group in collaboration with the International Breast Cancer Study Group. It included patients with high-risk early stage breast cancer who were randomly allocated to receive either high-dose chemotherapy with stem cell support more than 12 weeks or standard-dose chemotherapy more than 24 weeks [43]. The primary outcome measure of the trial was overall survival. Secondary outcome measures were quality-adjusted survival, disease-free survival, toxicity, HRQL, and cost-effectiveness. Eligible subjects were aged 16 to 65 years, and were about to start adjuvant chemotherapy for histologically confirmed early-stage primary breast cancer with five or more involved axillary nodes. Subjects were excluded if they were capable of only limited self-care and/or were confined to a bed or chair for more than 50% of waking hours (ECOG performance status 3 or 4). Enrolment was from March 1997 until March 2000 at multiple centers in Australia, New Zealand, Europe, and Asia. Subjects living in Australia and New Zealand were eligible to participate in a substudy. Substudy participants were required to provide detailed information about HRQL and resource usage by completing the UBQ-C and other questionnaires described in the following discussion. Questionnaires were completed before starting chemotherapy (baseline), 12 weeks after randomization (during chemotherapy), and a few months after completing chemotherapy.

Questionnaires and Other Characteristics of Subjects

The UBQ-C is a validated cancer-specific questionnaire that was designed as an outcome measure for clinical trials in the field of cancer. It includes 29 items about specific aspects of HRQL, and a single-item global scale that asks respondents to rate their global health status (health status thermometer) [23,44,45]. The 29 items about specific aspects of HRQL are grouped into subscales for physical function (three items), social/usual activities (four items), self-care (one item), and distresses (21 items) caused by physical and psychological symptoms relevant to cancer and its treatment. The UBQ-C also includes the general health item from the SF-36 health survey [46]. More details about the conceptual framework, development, composition, and psychometric properties of the UBQ-C are given in a companion paper [23].

Two additional questionnaires were completed. The Spitzer uniscale of global life quality was completed by all subjects as an additional global scale, but with the anchors of "highest quality" and "lowest quality" replaced by "best possible" and "worst possible" [47,48]. The Priestman and Baum linear analog self-assessment scales (LASAS) were completed by subjects in the advanced trial as validated measures of cancer-specific HRQL that include five scales about physical well-being, mood, pain, nausea and vomiting, and appetite [49,50]. Clinicians completed the ECOG performance status scale in the advanced trial. This rates patients' physical functional status as: "0"—fully active; "1"—restricted in physical activity but able to do light work; "2"—confined to a bed or chair for less than 50% of waking hours and capable of all self-care but unable to do any work; "3"—confined to a bed or chair for more than 50% of waking hours but capable of limited self-care; and "4"—totally confined to bed or chair, completely disabled, incapable of any self-care [51].

Statistical Methods

We optimized the scoring algorithm described in detail in a companion paper [23] and applied it to the clinical trial data sets. The scoring algorithm is outlined in Fig. 1.

First, we calculated subscale scores for physical function, social/usual activities, self-care, and distresses as the simple average of the nonmissing items, linearly transformed to a scale from 0 (worst) to 1 (best).

Indices were then calculated by applying the following formulae:

$$\text{Subset index} = [W1 \times PF] + [W2 \times SA] + [W3 \times SC] + [W4 \times DI] \quad (1)$$

$$W = \text{Var}(T) \times [1 - r(T)]/MSE(R) \quad (2)$$

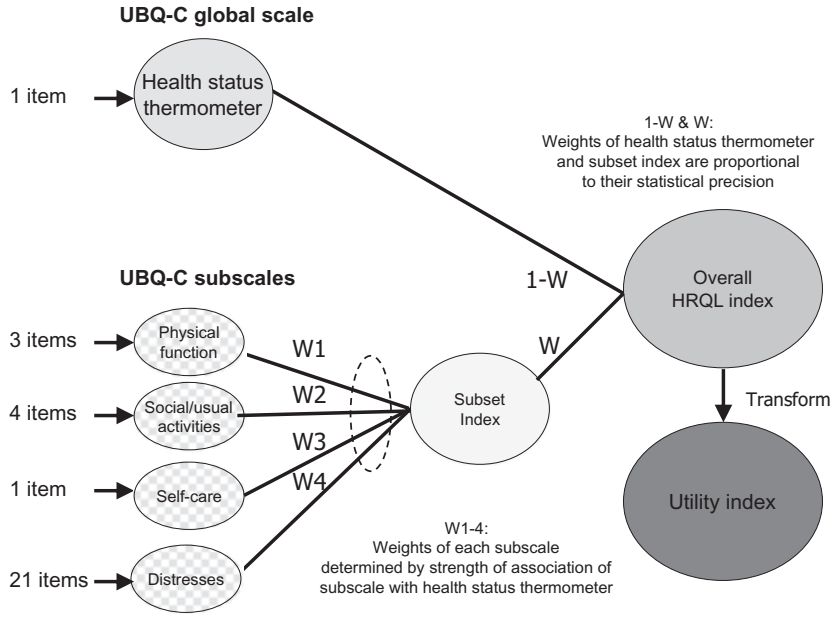

Figure 1 Deriving the overall HRQL index and utility index from the UBQ-C health-related quality of life questionnaire. HRQL, health-related quality of life; UBQ-C, Utility-Based Questionnaire-Cancer.

$$\text{Overall HRQL index} = [W \times \text{subset index}] + [(1 - W) \times HST] \quad (3)$$

$$\text{Utility index} = 1 - (1 - \text{overall HRQL index})^{2.03} \quad (4)$$

W1 to *W4* are the weights for the subscales, *PF* is physical function, *SA* is social/usual activities, *SC* is self-care, *DI* is distresses, and *HST* is the health status thermometer. *W* is the weight allocated to the subset index, so $1 - W$ is the weight allocated to the health status thermometer. Var(*T*) is the variance of the health status thermometer obtained from the data set. r(*T*) is the intraclass correlation coefficient of the health status thermometer, and was calculated with test–retest data from a previous validation study [45]. *MSE*(*R*) is the mean square for error from the linear regression of the health status thermometer on the four subscales, and was obtained from the data set.

Optimal weights for the subscales (*W1*–*W4*), subset index (*W*), and health status thermometer ($1 - W$) were derived for each trial using the ratings on the UBQ-C in the relevant data set. Weights *W*1 to *W*4 were derived from and proportional to the coefficients obtained from multivariable, ordinary least squares regression of the health status thermometer on the subscales. Weights *W* and ($1 - W$) were derived using Eq. 2.

The weights were then applied using Eqs. 1, 3, and 4 to calculate scores for the subset index, overall HRQL index, and utility index for each subject in each trial.

We examined the validity of the utility index against other characteristics of subjects. We tested its convergent validity, discriminative validity, responsiveness, and predictive validity by comparing it with other self-rated measures of HRQL and with measures of physical function, cancer stage, treatment phase, and subsequent survival.

Convergent validity tests how closely a measure is associated with related measures [52,53]. The convergent validity of the utility index was tested by Spearman rank correlation (r_S) with the Spitzer uniscale, the SF-36 general health item, and scales from the Priestman and Baum LASAS questionnaire referred to above. We expected substantial correlations with the Spitzer uniscale and the SF-36 general health item. Three clinical experts made a priori hypothesis about the expected values of r_S with the LASAS scales as: insignificant (<0.3), moderate (0.3–0.44), substantial (0.45–0.59), or high (>0.6). Hypotheses were considered supported by the data if the observed r_S were at least as high as the median of the experts' expected r_S.

Discriminative validity tests how well a measure can distinguish between groups defined by an alternate criterion [53,54]. The discriminative validity of the utility index was tested by its ability to detect cross-sectional differences between subjects with differing physical function as rated by their clinicians on the ECOG performance status scale referred to earlier. We also compared the discriminative ability of the UBQ-C overall HRQL index with that of the health status thermometer and the Spitzer-uniscale. Differences between groups were evaluated with Students *t* test.

Responsiveness tests the ability of a measure to detect clinically important change over time [54,55]. The responsiveness of the utility index was tested by comparing scores in the early cancer trial before, during, and after chemotherapy using paired *t* tests.

Predictive validity tests how closely a measure is associated with a subsequent outcome [52,53]. The predictive validity of the utility index was tested by its ability to predict survival duration in the advanced cancer trial, based on the hypothesis that overall survival in advanced cancer should be associated with baseline HRQL [56–60]. The strength of association between the utility index and survival duration was tested with the log-rank test, by dichotomizing subjects into a "poor HRQL" group (utility index less than or equal to the median) and a "good HRQL" group (utility index greater than the median).

The optimized scoring algorithm was applied to inform a specific treatment comparison of high-dose versus standard-dose chemotherapy for high-risk, early-stage breast cancer using the data collected during chemotherapy from the early cancer trial. First, we compared scores on the utility index for participants allocated to each treatment arm using unpaired *t* tests. Second, we used the index to reflect the relative importance of the effects of chemotherapy on different aspects of HRQL by comparing the weights allocated to each subscale. Third, we tested the hypothesis that the overall HRQL index compared with the health

status thermometer would give an estimate of the difference in mean scores between treatment groups that was more precise but unbiased. The relative precisions of the related measures were compared using a measure called the relative efficiency statistic [55,61]. The reciprocal of the relative efficiency statistic is the factor by which the sample size can be reduced when a more precise and therefore more efficient scale is used. The relative efficiency statistic was calculated as the squared ratio of the *t*-score for the index when comparing groups divided by the *t*-score for the related global measure when comparing groups.

Results

Study Profiles and Patient Characteristics

The study profiles describing the subjects in each trial are shown in Fig. 2. For the advanced cancer trial, compliance was excellent

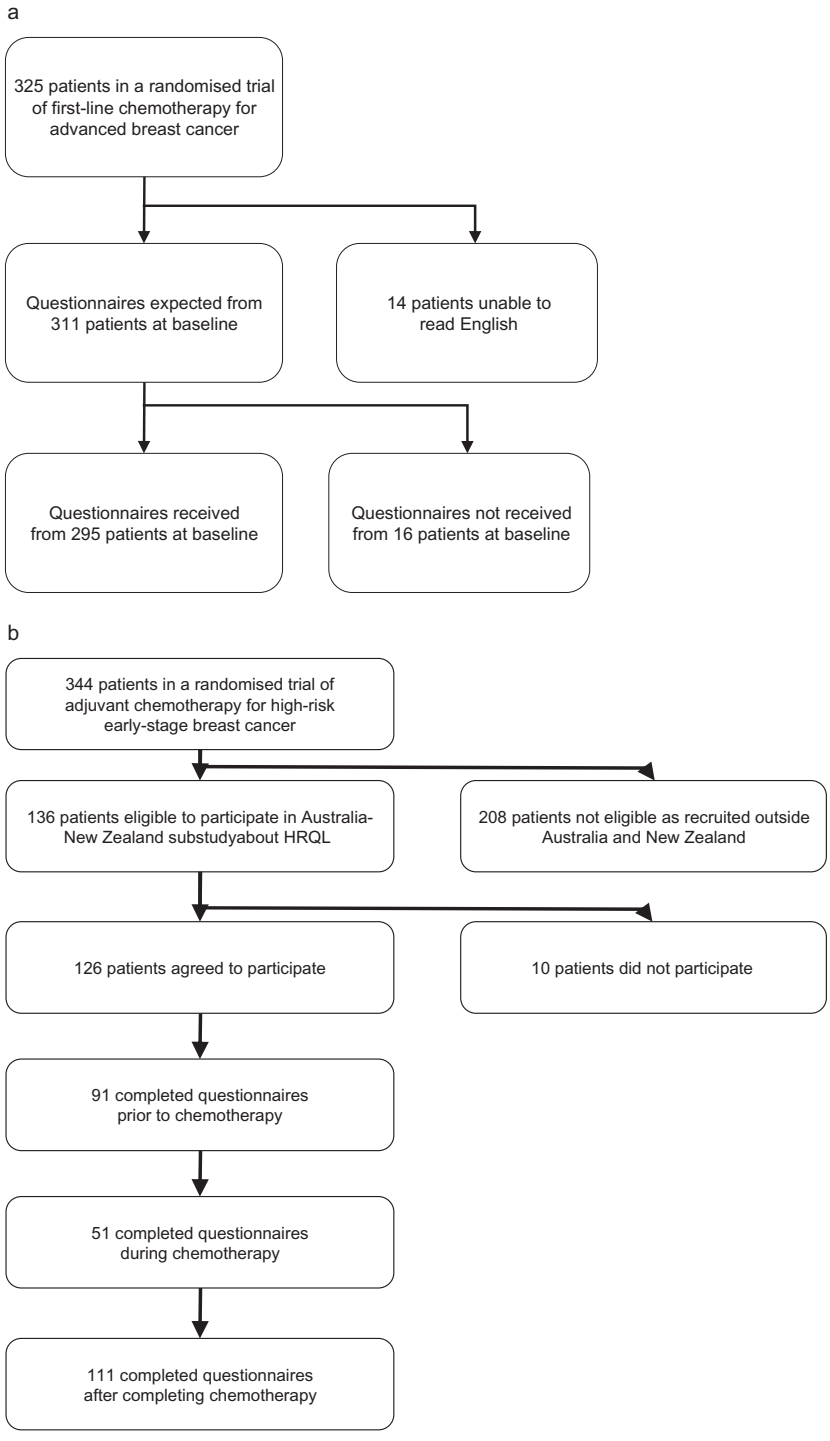

Figure 2 Study profile for (a) advanced and (b) early cancer trial.

Table 1 Patient characteristics

Data set	Advanced cancer trial (n = 325)	Early cancer trial (n = 126)
Cancer stage	Advanced	High-risk early-stage
Cancer type (%)		
Breast	100	100
Gender (%)		
Female	100	100
Age (years) (%)		
<40	2	14
40–49	12	47
50–59	29	35
60–69	36	3
≥70	21	—

Data set	Advanced cancer trial	Early cancer trial		
Treatment phase	Before treatment (n = 295)	Before treatment (n = 91)	During treatment (n = 51)	After treatment (n = 111)
General health (%)				
Excellent	6	22	6	22
Very good*	18	—	—	—
Good	30	54	40	66
Fair	32	19	42	9
Poor	13	4	12	3

Response category "Very good" not included in some versions of "General health" item of the Utility-Based Questionnaire-Cancer.

with questionnaires completed by 95% of subjects who were expected to complete them. For the early cancer trial, compliance was not as good with questionnaires completed by 72% before chemotherapy, 40% during chemotherapy, and 88% after completing it. All items on each UBQ-C questionnaire except for "Sex life" and "Other problems" were completed by more than 90% of subjects in both trials. Characteristics of the 421 patients are shown in Table 1. Data was obtained from patients with breast cancer of both early and advanced stages. All subjects were female and most age groups were represented. For the advanced cancer trial, most had good performance status (ECOG 0 in 34% and ECOG 1 in 54%), and fewer had poor performance status (ECOG 2 in 11% and ECOG 3 in 2%). Ratings of general health ranged from "Excellent" to "Poor."

Subjects' ratings on the UBQ-C are summarized in Table 2. At baseline, patients with advanced cancer reported worse health status than patients with early cancer as expected [means of 0.69 vs. 0.81, difference 0.13 (with rounding), 95% CI 0.08 to 0.17, $P < 0.0001$]. Patients with early cancer reported worse health status during chemotherapy than before starting it [means 0.68 vs. 0.81, mean deterioration 0.13, 95% confidence internal (CI) 0.08 to 0.19, $P < 0.0001$]; or after finishing it [means 0.68 vs. 0.84, mean improvement 0.15 (with rounding), 95% CI 0.10 to 0.21, $P < 0.0001$]. Similar differences were reported for ratings on UBQ-C subscales (Table 2).

Optimized Scoring Algorithms

The optimized index weights for the subset index (W), health status thermometer (1-W), and subscales (W1-4) for each trial are shown in Table 3. The weight assigned to the health status thermometer was similar for each trial and accounts for about two-thirds of the overall HRQL index. Of the subscales, greatest weight was given to distresses and least to self-care. The ordering of the weights assigned to the advanced cancer trial and early cancer trial were similar. Distresses were assigned the greatest weight, followed by social/usual activities, physical function, and self-care. However greater weight was assigned to distresses, and less weight to physical function and self-care, in women with early breast cancer than in women with advanced cancer.

Validation

Comparisons of the utility index with other characteristics of subjects supported its validity.

The convergent validity of the utility index was supported by its substantial correlation with the SF-36 general health item in both trials (r_S 0.74 in advanced and 0.64 in early) and the Spitzer uniscale in advanced cancer (r_S 0.71). There was also complete concordance of all expected and observed correlations of the utility index with the Priestman and Baum LASAS in the advanced cancer trial (data not shown).

The discriminative validity of the utility index was supported by strong evidence that subjects with early breast cancer before starting chemotherapy had higher utilities than those with advanced breast cancer [mean difference 0.07 (with rounding), 95% CI 0.04 to 0.10, $P < 0.0001$] (Table 2). The discriminative validity of the utility index was also supported by its ability to distinguish subjects with differing performance status (PS) as rated by their clinicians in the advanced cancer trial (good performance status: mean 0.90, 95% CI 0.88 to 0.91; poor performance status: mean 0.73, 95% CI 0.67 to 0.79; mean difference 0.17, 95% CI 0.12 to 0.21; $P < 0.0001$).

The responsiveness of the utility index was supported by strong evidence that subjects with early breast cancer had higher utilities before starting chemotherapy than during it (mean difference 0.07, 95% CI 0.04 to 0.10; $P < 0.0001$) (Table 2).

The predictive validity of the utility index was supported by its ability to predict survival duration in the advanced cancer trial

Table 2 Ratings on the health status thermometer, UBQ-C subscales, overall HRQL index, and utility index

Data set	Advanced cancer trial		Early cancer trial					
	Before treatment		Before treatment		During treatment		After treatment	
Treatment phase	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Health status thermometer	0.69	0.20	0.81	0.15	0.68	0.21	0.84	0.13
UBQ-C subscales								
Physical function	0.53	0.32	0.77	0.21	0.63	0.24	0.80	0.20
Social/usual activities	0.66	0.29	0.74	0.23	0.69	0.22	0.88	0.17
Self-care	0.89	0.20	0.89	0.15	0.97	0.10	0.99	0.06
Distresses	0.78	0.15	0.77	0.15	0.69	0.18	0.83	0.13
Overall HRQL index	0.69	0.18	0.80	0.13	0.68	0.18	0.84	0.12
Utility index	0.88	0.13	0.94	0.07	0.87	0.15	0.96	0.06

All ratings on scale from best (one) to worst (zero).
HRQL, health-related quality of life; SD, standard deviation; UBQ-C, Utility-Based Questionnaire-Cancer.

Table 3 Weights for the health status thermometer and subscales

		Advanced cancer trial	Early cancer trial
W1	Physical function	0.20	0.09
W2	Social/usual activities	0.23	0.25
W3	Self-care	0.04	0.01
W4	Distresses	0.53	0.64
1-W	Health status thermometer	0.66	0.63
W	Subset index	0.34	0.37

1-W, W, W1-4 refer to the weights assigned to the health status thermometer, subset index, and subscales in Eqs. 1 and 2 (see discussion and Figure 1).

when patients were divided into roughly equal-sized groups above and below the median score on the utility index (Fig. 3). There was strong evidence that subjects with worse scores on the utility index at baseline (<0.92) had shorter survival than those with higher scores (median 17 vs. 23 months, log-rank $P = 0.005$).

Treatment Comparison

The scoring algorithm was applied to the treatment comparison of high-dose chemotherapy versus standard-dose chemotherapy for early-stage breast cancer. Subjects receiving high-dose chemotherapy reported worse impairment of most specific aspects of HRQL (Fig. 4), which was expected because high-dose chemotherapy is more toxic in this setting [43]. There was a trend to better mean scores on the utility index for patients allocated to standard-dose chemotherapy (mean 0.95) compared with high-dose chemotherapy (mean 0.92) with mean difference of –0.03 (95% CI –0.07 to 0.01, $P = 0.10$). The overall HRQL index gave stronger evidence of this effect (mean difference –0.07, 95% CI –0.13 to –0.01, $t = 2.36$, $P = 0.02$) than the health status thermometer (mean difference –0.06, 95% CI –0.12 to 0.01, $t = 1.72$, $P = 0.09$) (Fig. 4). The relative efficiency of the overall HRQL index compared with the health status thermometer was 1.9. In this practical illustration, the improvement in precision by using the overall HRQL index compared with the health status thermometer was sufficient to conclude that the more toxic regimen causes significantly worse effects on overall HRQL.

Conclusions

We have applied a scoring algorithm for a cancer-specific utility-based instrument to clinical trial data sets and illustrated how it can be optimized in different clinical contexts. The algorithm converts ratings from a cancer-specific questionnaire for HRQL into a utility index that is based on the perspective of cancer patients. First, we optimized the scoring algorithm in two different clinical contexts for breast cancer by adjusting the index weights using data from two clinical trials. Second, we applied the algorithm to generate utility scores. Third, we showed that the utility index had convergent validity with related scales from other instruments, discriminative validity between participants with differing performance status, responsiveness to toxic effects of chemotherapy in early cancer, and predictive validity about subsequent survival duration. Fourth, we used the utility index to inform a treatment comparison of high-dose chemotherapy with stem-cell support versus standard-dose chemotherapy for high-risk, early-stage breast cancer. It can be used to generate sensitive and responsive utility scores, and quality-adjusted life-years that can be used within a trial to compare the net benefit of treatments and inform clinical decision-making.

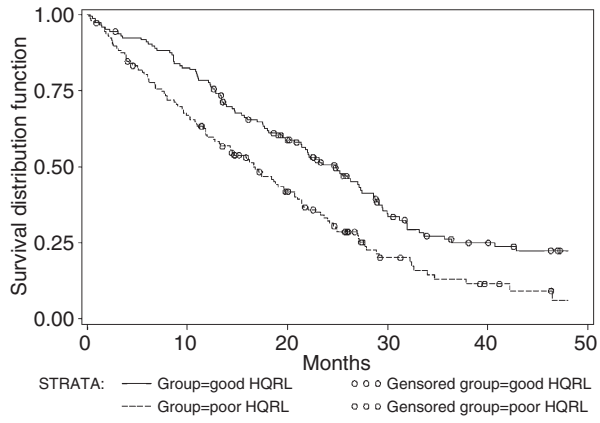

Figure 3 Kaplan–Meier plots for survival duration of subjects grouped by utility index in advanced cancer trial. Good health-related quality of life (HRQL), score on utility index <0.92. Poor HRQL, score on utility index ≥0.92.

The novelty of the approach described in this paper is that the scoring algorithm can be optimized for different clinical contexts. In contrast, most scoring algorithms for utility-based instruments use the same scoring algorithm across different diseases and treatments [1,17,20–22]. The algorithm is optimized by giving additional weight to the subscales about specific aspects of HRQL that are most closely associated with a single-item global scale (the health status thermometer) in the relevant data set. The reason to optimize the algorithm in different contexts is to reflect variations in patients' attitudes, preferences, and priorities across different cancer types, stages, and treatments [38,39].

The primary benefit of optimizing the scoring algorithm for each clinical context is that it should better reflect the perspective of the individuals in that situation. For example, in the comparison of high-dose versus standard-dose chemotherapy for early-stage breast cancer (Fig. 3), there were large differences in distresses and physical function but little or no difference in self-care. Combinations of the subscales giving greater weight to self-care would yield little difference between high-dose and standard-dose chemotherapy, whereas those giving greater weight to distresses and physical function would favor standard-dose chemotherapy. We assigned weight according to correlations with the health status thermometer, resulting in significant differences between treatments on the indices for overall HRQL and utility that should reflect the preferences and attitudes of the women in the trial.

Optimizing the scoring algorithm could give more precise estimates of clinically important differences in utility between patient groups, because the index is focussed on those aspects of HRQL that are most relevant to those patients. A more precise utility index will reduce the uncertainty around the incremental effectiveness of treatments in sensitivity analyses, because it is more responsive to small but meaningful effects of cancer treatments [62]. A more precise utility index will also reduce the sample size required to detect a given difference with a given level of precision [62].

Another benefit of optimizing the scoring algorithm for each clinical context is that the ordering of the weights can inform clinicians and researchers about the importance of various symptoms, side effects, and dysfunctions that patients in different clinical contexts most wish to avoid. For example, in both data sets we found that greatest weight was given to distresses, followed by social/usual activities, physical function, and self-care (Table 3). The ordering of the weights assigned to each subscale may be related to several factors. The large weight assigned to distresses may reflect the emotional distress that most patients experienced caused by having cancer, the physical symptoms of

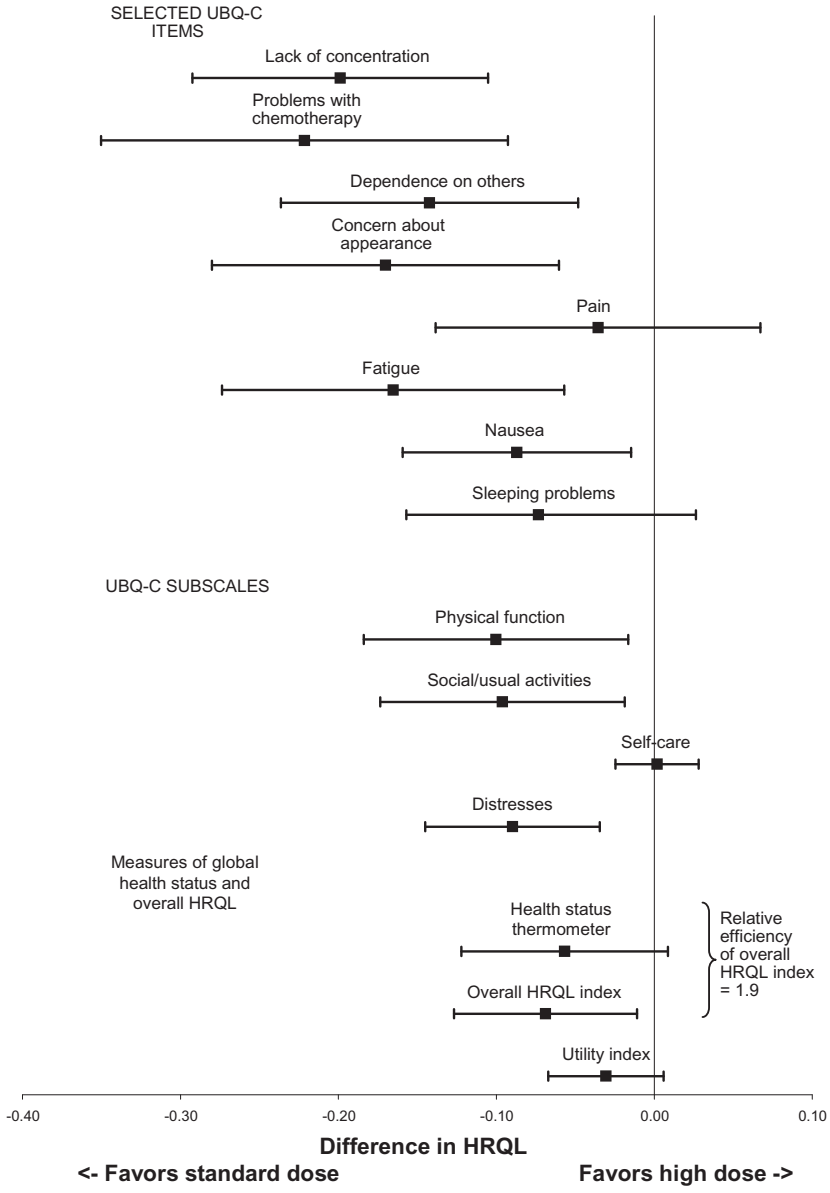

Figure 4 Differences in HRQL between treatment groups for subjects during treatment in early cancer trial, based on: 1) selected UBQ-C items; 2) UBQ-C subscales; 3) health status thermometer; 4) overall HRQL index; and 5) utility index. All ratings on scale from 0 to 1. HRQL, health-related quality of life.

advanced cancer, and the side effects of toxic chemotherapy for early-stage cancer. The low weight assigned to self-care may reflect the lack of problems with self-care that most patients reported in each trial. The ordering of the weights may also reflect the number of items within each subscale, with distresses (21 items) assigned greater weight than physical function (three items), social/usual activities (four items), or self-care (one item). There were some differences in weights between data sets. Greater weight was assigned to distresses and less weight was assigned to physical function for early cancer compared with advanced cancer (Table 3). The greater weight assigned to distresses for patients with early cancer may reflect their greater emotional distress caused by a recent diagnosis of cancer, and their experience of side effects from chemotherapy that had not yet been administered to the patients with advanced cancer. The lower weight assigned to physical function for patients with early cancer is probably explained by the absence of the deterioration in physical function that occurs with advanced cancer. This information can be used by researchers to design more targeted interventions to improve HRQL in the dimensions of greatest importance to patients, and by all health-care workers to improve counselling of patients [39].

We recommend that the scoring algorithm is optimized for each clinical context in which it is used. This is a potential limitation in that it requires additional analyses and familiarity with Lumley's method. Another limitation is that the utility scores may not be comparable from one disease or treatment context to another, because the scoring algorithm and its index weights cannot be standardized across trials [38]. Consequently we recommend that our approach is used to compare treatments in the context of a trial in a well-defined population for a specific clinical condition, because the attitudes of patients are likely to be more similar. It is less suited to studies that include diverse populations, or for comparing utilities and quality-adjusted life-years from one study or context to another, because the attitudes of patients will be more diverse. Comparability of utility scores is

a key requirement when utilities are used to inform economic decisions, because health funders and policy makers make decisions across diseases and contexts [1,63]. However, comparability of utility scores is less important when utilities are used to inform clinical decisions, because a clinical decision is always limited to a single disease type and stage. The requirements of utility scores used to inform clinical decisions are that they reflect the experiences of the patients under study, and are valid, sensitive, and reliable.

The measurement properties of the utility index reported in this paper support its validity as a measure of HRQL for the clinical context of chemotherapy for early and advanced breast cancer. The utility index had convergent validity with independent scales of general health and global quality of life, was able to discriminate patients with different stages of cancer, and was responsive to changes attributable to having chemotherapy. Another way to validate a utility index is to compare the scores derived by the utility index with utilities elicited directly from the same patients with a time trade-off interview. This could be performed in future studies.

Future research is also needed to determine if optimization of the scoring algorithm for each context makes a meaningful difference to the utility scores and QALYs generated from the utility index, their sensitivity and responsiveness to detect differences between treatment groups, and most importantly to the outcome of clinical decisions in specific clinical contexts.

Finally, it is important to comment on the strengths and limitations of the data sets used in this study. Patients participating in a clinical trial of treatments are the ideal source of information about the effects of those treatments on HRQL. The data sets included patients with early and advanced cancer, before, during, and after chemotherapy. Compliance was good with the questionnaire completion, particularly for the advanced cancer trial. We used validated cancer-specific questionnaires that included a broad range of items about specific aspects of HRQL that are commonly affected by cancer and side effects of treatment. A limitation of the data sets is that they only included women in Australia and New Zealand with breast cancer receiving chemotherapy, so the results may not be applicable to other cancer types or treatments, other countries, and men. Further application and validation of the utility index is ongoing in other clinical contexts including chemotherapy for advanced colorectal cancer and hormonal therapy for the prevention of breast cancer [64,65]. The colorectal study includes British and male subjects. Compliance with completing questionnaires in the early cancer trial was poor during chemotherapy. Patients who do not complete questionnaires tend to have worse HRQL [66], so the analyses may underestimate the detrimental effects of treatment on HRQL. Finally, the early cancer trial is relatively old so the effects of chemotherapy may be different to that with more modern treatments. Therefore the utility scores generated from it may not be appropriate for informing clinical or health policy decisions about current treatments.

Our approach enables HRQL data obtained with a simple questionnaire to be converted into utility scores by using an optimized scoring algorithm that reflects the perspective of the cancer patients under study. The approach is flexible and applicable to other trials and other HRQL instruments. Generation of utility scores based on HRQL data collected within a clinical trial provides an ideal source of information to inform clinical decisions, and to add a useful additional perspective to inform health policy and economic decisions.

Source of financial support: Peter Grimison was supported by National Health and Medical Research Council, Australia; Cancer Institute NSW, Australia; and Glaxo Smith Kline Australia. The funding agencies had no role in the study design, the collection or analysis of the data, the interpretation of the results, the preparation of the manuscript, or the decision to submit the manuscript for publication.

References

1 Drummond MF. Methods for the Economic Evaluation of Health Care Programmes (3rd ed.). Oxford: Oxford University Press, 2005.
2 Glasziou PP, Cole BF, Gelber RD, et al. Quality adjusted survival analysis with repeated quality of life measures. Stat Med 1998;17:1215–29.
3 Stiggelbout AM, de Haes JCJM. Patient preference for cancer therapy: an overview of measurement approaches. J Clin Oncol 2001;19:220–30.
4 Au H-J, Golmohammadi K, Younis T, et al. Cost-effectiveness analysis of adjuvant docetaxel, doxorubicin, and cyclophosphamide (TAC) for node-positive breast cancer: modeling the downstream effects. Breast Cancer Res Treatt (Online First) 2008; doi 10.1007/s10549-008-0034-1.
5 Takeda AL, Jones J, Loveman E, et al. The clinical effectiveness and cost-effectiveness of gemcitabine for metastatic breast cancer: a systematic review and economic evaluation. Health Technol Assess 2007;11:1–80.
6 Locker GY, Mansel R, Cella D, et al. Cost-effectiveness analysis of anastrozole versus tamoxifen as primary adjuvant therapy for postmenopausal women with early breast cancer: a US healthcare system perspective. The 5-year completed treatment analysis of the ATAC ("Arimidex", Tamoxifen Alone or in Combination) trial. Breast Cancer Res Treat 2007;106:229–38.
7 Kurian AW, Thompson RN, Gaw AF, et al. A cost-effectiveness analysis of adjuvant trastuzumab regimens in early her2/neu-positive breast cancer. J Clin Oncol 2007;25:634–41.
8 Brink M, van den Hout WB, Stiggelbout AM, et al. Cost-utility analysis of preoperative radiotherapy in patients with rectal cancer undergoing total mesorectal excision: a study of the Dutch colorectal cancer group. J Clin Oncol 2004;22:244–53.
9 Eckermann SD, Martin AJ, Stockler MR, Simes RJ. The benefits and costs of tamoxifen for breast cancer prevention. Aust N Z J Public Health 2003;27:34–40.
10 Remak E, Charbonneau C, Negrier S, et al. Economic evaluation of sunitinib malate for the first-line treatment of metastatic renal cell carcinoma. J Clin Oncol 2008;26:3995–4000.
11 Mallick R, Hudes G, Levy DE. Clinically important differences in quality-adjusted survival in patients with advanced renal cell carcinoma receiving first-line treatment with temsirolimus or interferon-alpha (Abstract PCN64). Value Health 2007;10:A342.
12 Bernhard J, Zahrieh D, Zhang JJ, et al. Quality of life and quality-adjusted survival (Q-TWiST) in patients receiving dose-intensive or standard dose chemotherapy for high-risk primary breast cancer. Br J Cancer 2008;98:25–33.
13 Bernhard J, Zahrieh D, Coates AS, et al. Quantifying trade-offs: quality of life and quality-adjusted survival in a randomised trial of chemotherapy in postmenopausal patients with lymph node-negative breast cancer. Br J Cancer 2004;91:1893–901.
14 Elkin EB, Cowen ME, Cahill D, et al. Preference assessment method affects decision-analytic recommendations: a prostate cancer treatment example. Med Decis Making 2004;24:504–10.
15 Grann VR, Jacobson JS, Thomason D, et al. Effect of prevention strategies on survival and quality-adjusted survival of women with brca1/2 mutations: an updated decision analysis. J Clin Oncol 2002;20:2520–9.
16 Sommers BD, Beard CJ, D'Amico AV, et al. Decision analysis using individual patient preferences to determine optimal treatment for localized prostate cancer. Cancer 2007;110:2210–17.
17 Green C, Brazier J, Deverill M. Valuing health-related quality of life. A review of health state valuation techniques. Pharmacoeconomics 2000;17:151–65.
18 Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. J Clin Epidemiol 2003;56:317–25.

19 Froberg DG, Kane RL. Methodology for measuring health-state preferences–I: measurement strategies. J Clin Epidemiol 1989;42:345–54.
20 Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
21 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.
22 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.
23 Grimison P, Simes RJ, Hudson HM, Stockler MR. Deriving a patient-based utility index from a cancer-specific quality of life questionnaire. Value Health 2008 (in press).
24 Brazier J, Czoski-Murray C, Roberts J, et al. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. Med Decis Making 2008;28:113–26.
25 Dobrez D, Cella D, Pickard AS, et al. Estimation of patient preference-based utility weights from the functional assessment of cancer therapy-general. Value Health 2007;10:266–72.
26 Berger ML, Bingefors K, Hedblom EC, et al. Health Care Cost, Quality, and Outcomes: ISPOR Book of Terms. Lawrenceville, NJ: International Society for Pharmacoeconomics and Outcomes Research, 2003.
27 Noyes K, Dick AW, Holloway RG. The implications of using us-specific EQ-5D preference weights for cost-effectiveness evaluation. Med Decis Making 2007;27:327–34.
28 Australian Government Department of Health and Ageing. Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.0). Canberra, Australia: Australian Government Department of Health and Ageing, 2006.
29 National Institute for Clinical Excellence (NICE). Guide to the Methods of Technology Appraisal. London: NICE, 2004.
30 Russell LB, Gold MR, Siegel JE, et al. The role of cost-effectiveness analysis in health and medicine. Panel on cost-effectiveness in health and medicine. JAMA 1996;276:1172–7.
31 Gold MR. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.
32 Stiggelbout AM, de Vogel-Voogt E. Health state utilities: a framework for studying the gap between the imagined and the real. Value Health 2008;11:76–87.
33 Fryback DG. A US valuation of the EQ-5D. Med Care 2005;43:199.
34 Johnson JA, Luo N, Shaw JW, et al. Valuations of EQ-5D health states: are the United States and United Kingdom different? Med Care 2005;43:221.
35 Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care 2005;43:203–20.
36 Glasziou P, Alexander J, Beller E, Clarke P. Which health-related quality of life score? A comparison of alternative utility measures in patients with type 2 diabetes in the ADVANCE trial. Health Qual Life Outcomes 2007;5:21.
37 Huang IC, Willke R, Atkinson M, et al. US and UK versions of the EQ-5D preference weights: does choice of preference weights make a difference? Qual Life Res 2007;16:1065–72.
38 Lumley T, Simes RJ, Gebski V, Hudson HM. Combining components of quality of life to increase precision and evaluate trade-offs. Stat Med 2001;20:3231–49.
39 Osoba D, Hsu M-A, Copley-Merriman C, et al. Stated preferences of patients with cancer for health-related quality-of-life (HRQOL) domains during treatment. Qual Life Res 2006; 15:273–83.
40 Rose M, Scholler G, Klapp BP, Bernheim JL. Weighting dimensions in "generic" QOL questionnaires by Anamnestic Comparative Self-assessment: different weights in different diseases (Abstract). Qual Life Res 1998;7:655.
41 Wittenberg E, Halpern E, Divi N, et al. The effect of age, race and gender on preference scores for hypothetical health states. Qual Life Res 2006;15:645–53.
42 Stockler M, Sourjina T, Grimison P, et al. A randomized trial of capecitabine (C) given intermittently (IC) versus continuously (CC) versus classical CMF as first line chemotherapy for advanced breast cancer (ABC). J Clin Oncol, ASCO Annual Meeting Proceedings (Post-Meeting Edition). 2007;25: 1031.
43 International Breast Cancer Group. Multicycle dose-intensives chemotherapy for women with high-risk primary breast cancer: results of international breast cancer study group trial 15-95. J Clin Oncol 2006;24:370–8.
44 Martin AJ, Greatorex V, Simes RJ. Preliminary results on the reliability and validity of the UBQ-C(ancer) items. Technical Report No 9601: NHMRC Clinical Trials Centre, University of Sydney, 1996.
45 Martin AJ, Greatorex V, Simes RJ. Towards a utility-based assessment for cancer patients: Reliability and validity of the UBQ-C(ancer) items (Abstract P81). Control Clin Trials 1997;18(Suppl):S160–60.
46 Ware JE, Jr., Sherbourne, CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473–83.
47 Sloan JA, Loprinzi CL, Kuross SA, et al. Randomized comparison of four tools measuring overall quality of life in patients with advanced cancer. J Clin Oncol 1998;16:3662–73.
48 Spitzer WO, Dobson AJ, Hall J, et al. Measuring the quality of life of cancer patients: a concise QL-Index for use by physicians. J Chronic Dis 1981;34:585–97.
49 Coates A, Gebski V, Signorini D, et al. Prognostic value of quality-of-life scores during chemotherapy for advanced breast cancer. Australian New Zealand Breast Cancer Trials Group. J Clin Oncol 1992;10:1833–38.
50 Priestman TJ, Baum M. Evaluation of quality of life in patients receiving treatment for advanced breast cancer. Lancet 1976;307:899–900.
51 Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the eastern cooperative oncology group. Am J Clin Oncol 1982;5:649–55.
52 Fayers P, Machin D. Quality of Life: The Assessment, Analysis and Interpretation of Patient-Reported Outcomes (2nd ed.). Chichester, England: John Wiley & Sons, 2007.
53 Stockler MR, Osoba D, Corey P, et al. Convergent discriminitive, and predictive validity of the prostate cancer specific quality of life instrument (PROSQOLI) assessment and comparison with analogous scales from the EORTC QLQ-C30 and a trial-specific module. J Clin Epidemiol 1999;52:653–66.
54 Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. (3rd ed.). Oxford, New York: Oxford University Press, 2003.
55 Stockler MR, Osoba D, Goodwin P, et al. Responsiveness to change in health-related quality of life in a randomized clinical trial: a comparison of the prostate cancer specific quality of life instrument (PROSQOLI) with analogous scales from the EORTC QLQ-C30 and a trial specific module. J Clinical Epidemiol 1998;51:137–45.
56 Bottomley A, Efficace F. Predicting survival in advanced cancer patients: is it possible with patient-reported health status data? Ann Oncol 2006;17:1037–8.
57 Coates AS, Hurny C, Peterson HF, et al. Quality-of-life scores predict outcome in metastatic but not early breast cancer. J Clin Oncol 2000;18:3768–74.
58 Dancey J, Zee B, Osoba D, et al. Quality of life scores: an independent prognostic variable in a general population of cancer patients receiving chemotherapy. The national cancer institute of Canada clinical trials group. Qual Life Res 1997;6:151–8.
59 Siddiqui F, Kachnic LA, Movsas B. Quality-of-life outcomes in oncology. Hematol Oncol Clin North Am 2006;20:165–85.
60 Gotay CC, Kawamoto CT, Bottomley A, et al. The Prognostic significance of patient-reported outcomes in cancer clinical trials. J Clin Oncol 2008;26:1355–63.
61 Armitage P, Berry G, Matthews JNS. Statistical Methods in Medical Research. Massachusetts: Blackwell Publishing, 2002.

62 McKenna SP, Ratcliffe J, Meads DM, Brazier JE. Development and validation of a preference based measure derived from the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) for use in cost utility analyses. Health Qual Life Outcomes 2008;6:65.
63 Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford: Oxford University Press, 2007.
64 Grimison PS, Coates AS, Forbes JF, et al. Tamoxifen (TAM) for the prevention of breast cancer: importance of specific aspects of health-related quality of life (HRQL) to global health status in the ANZ BCTG substudy of IBIS-1 (ANZ 92P1) (Abstract 1516). J Clin Oncol ASCO Annual Meeting Proceedings 2008;25:15S–1516.
65 Stockler MR, Grimison PS, Price TJ, et al. Comparing utilities for advanced colorectal cancer valued from societal and cancer-patients' perspectives using baseline data from the MAX study (Abstract 6504). J Clin Oncol ASCO Annual Meeting Proceedings 2008;25:15S–6504.
66 Bernhard J, Cella DF, Coates AS, et al. Missing quality of life data in cancer clinical trials: serious problems and challenges. Stat Med 1998;17:517–32.