Deriving a Patient-Based Utility Index from a Cancer-Specific Quality of Life Questionnaire

Peter S. Grimison, BSc(Med), MBBS(Hons), MPH-FRACP,1 R. John Simes, BSc(Med)(Hons), MBBS(Hons), MD, SM, FRACP,1 H. Malcolm Hudson, BSc(Hons), PhD,1 Martin R. Stockler, MBBS(Hons), MSc, FRACP1,2

1NHMRC Clinical Trials Centre, University of Sydney, Australia; 2Sydney Cancer Centre, Royal Prince Alfred Hospital, Camperdown, Australia

ABSTRACT

Objectives: The aim of this study was to derive a scoring algorithm for a validated disease-specific quality of life instrument called the Utility-Based Questionnaire-Cancer (UBQ-C) that provided a utility index designed to inform clinical decisions about cancer treatments.

Methods: The UBQ-C includes a scale for global health status (1 item); and subscales for physical function (3 items), social/usual activities (4 items), self-care (1 item), and distresses because of physical and psychological symptoms (21 items). A scoring algorithm was derived to convert the subscales into a subset index, and combine it with the global scale into an overall health-related quality of life (HRQL) index, which was converted to a utility index with a power transformation. The valuation survey consisted of 204 advanced cancer patients who completed the UBQ-C and assigned time trade-off (TTO) utilities about their own health state. Preliminary validation involved comparing these derived utilities with other measures of HRQL.

Results: Weights for the subset index were: physical function 0.28, social/usual activities 0.06, self-care 0.01, and distresses 0.64. Weights for the overall HRQL index were health status 0.65 and subset index 0.35. The mean of the utility index scores was similar to the mean of the TTO utilities (0.92 vs. 0.91, P = 0.6). The utility index was substantially correlated with other measures of HRQL.

Conclusions: Data from a simple, self-rated, disease-specific questionnaire can be converted into a utility index suitable for comparing the net effect of cancer treatments on quality of life, and to evaluate trade-offs between quality and quantity of life in quality-adjusted survival analyses.

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

Introduction

Utility-based instruments are a common means of generating utility scores for calculating quality-adjusted life-years (QALYs) [1]. A utility-based instrument generally consists of a questionnaire that elicits responses about multiple dimensions of health status and health-related quality of life (HRQL), and a scoring algorithm that is used to convert the ratings on the questionnaire into a single utility-based index [1,2]. The scoring algorithms for utility-based instruments are valued in surveys, where subjects are asked to assign utilities to the health states defined by the questionnaire [1,2]. For example, the valuation survey for the EQ-5D instrument involved lay people assigning utilities with the time trade-off (TTO) method to a number of hypothetical health states defined in five generic dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) [3,4]. Utility-based instruments may vary in the type of questions they contain (generic or disease-specific), and the perspective from which the scoring algorithm is valued (lay people or patient) [1,5,6].

Controversy exists about the suitability of disease-specific utility-based instruments for generating utility scores. Generic instruments like the EuroQol EQ-5D [3], Health Utilities Index (HUI3) [7], or SF-6D [8] ask about core aspects of HRQL that are of interest in a wide range of settings. The main argument for using a generic utility-based instrument is that it allows comparisons across a wide range of diseases and healthy populations [9–11]. However, a generic instrument is likely to provide an inadequate description of many diseases, so the utility scores that it generates may be insensitive to differences between individuals with that disease [9,12–14]. Disease-specific, utility-based instruments were designed to address this lack of sensitivity by asking about specific aspects of HRQL relevant to that disease or condition [6,12,15,16].

Controversy also exists about the utility scores that are valued from the perspective of lay people versus patients [10,17,18]. The distinction is important because a patient typically assigns a higher utility to a health state than a lay person [18–20]. Economic guidelines generally recommend the use of generic utility-based instruments based on the perspective of lay people [21–23]. The main argument for using the perspective of lay people for informing funding and policy decisions is that the primary objective in a publicly funded health system is to maximize health for society [1]. It is generally recommended that the perspective of patients is used to inform clinical decision-making [1,5,24,25]. The main argument for using the perspective of patients for clinical decisions is that the primary objective is to maximize health for the individual patient experiencing that condition [26,27].

We posit that disease-specific instruments valued by patients are preferable for informing clinical decisions, whereas generic instruments valued by lay people may be preferable for decisions about health policy.

The aim of this study was to derive a scoring algorithm for a disease-specific, utility-based, HRQL instrument that is designed to inform clinical decisions about cancer treatments. The algorithm converts ratings from a cancer-specific HRQL questionnaire into a utility-based index designed to reflect the perspective of cancer patients. This article describes the development and preliminary validation of the algorithm. A companion article describes the application of the algorithm to trial datasets, and illustrates how it can be optimized in different treatment contexts [28].

Address correspondence to: Peter Grimison, NHMRC Clinical Trials Centre, University of Sydney, Locked Bag 77, Camperdown, Australia. E-mail: Peter.Grimison@sswahs.nsw.gov.au

10.1111/j.1524-4733.2009.00505.x
Methods

Source of Data

The valuation survey used to derive the scoring algorithm involved ambulatory patients with advanced cancer and impaired HRQL who were recruited from two tertiary-referral oncology outpatient units [29]. Eligible patients had advanced cancer, impaired HRQL, and were willing and able to complete a self-administered HRQL questionnaire and participate in a 1-hour interview in English. All patients provided written informed consent. The study was approved by the human research ethics committees at all participating institutions.

Consenting patients were registered and scheduled for an interview, usually on the day of their next appointment at the oncology clinic. Utilities were elicited directly from subjects about their current HRQL by one trained researcher using a standardized, face-to-face, TTO interview with a hypothetical survival time of 2 years. The TTO was expressed on the standard continuum where 1 represents full health and 0 represents dead [24]. Patients were mailed the questionnaire and asked to complete it 3 to 7 days before the planned interview.

The Utility-Based Questionnaire-Cancer

The Utility-Based Questionnaire is a validated, disease-specific HRQL questionnaire that was designed to be an outcome measure for clinical trials in cancer and cardiovascular disease [29–31]. The cancer version (the Utility-Based Questionnaire-Cancer [UBQ-C]) includes 29 items about specific aspects of HRQL, and a global scale called the health status thermometer, which is a single item that asks respondents for a unified assessment of their health status. The 29 items about specific aspects of HRQL are grouped into subscales for physical function (3 items), social/usual activities (4 items), self-care (1 item), and distresses (21 items) because of physical and psychological symptoms associated with cancer and its treatment.

The UBQ-C was designed for use in clinical trials of cancer therapy, so it needed to be relevant to cancer patients, relatively brief and easy to self-complete. The form and content of the questionnaire build on the conceptual framework for health status assessment developed by Gudex and Kind for an existing generic preference-based instrument called the Health Measurement Questionnaire (HMQ) [32]. The HMQ includes 36 items covering five key dimensions of HRQL (general mobility, usual activities, self-care activities, social and personal relationships, and psychological distresses). A generic core set of items from the HMQ was taken for the UBQ-C. Items less relevant for cancer patients (e.g., hearing, vision, writing, speaking, and incontinence) were discarded, and the response formats of some items were modified. Cancer-specific items were selected for addition by a review of existing literature on HRQL instruments used in cancer patients [33,34]. Two measures of global health status were also added. The health status thermometer is similar to the graduated, vertical, visual analog scale that accompanies the EuroQol EQ-5D questionnaire [35,36], but with the anchors of “best imaginable health state” and “worst imaginable health state” replaced by “full health” and “death,” to conform with the requirements of a utility scale. The UBQ-C also includes the general health item from the Short-Form-36 health survey (SF-36) [37], which is a widely used and extensively validated measure of generic health status [38]. The version of this item used in the valuation survey described above omitted the response category “very good.”

The psychometric properties of the UBQ-C in a cancer population have been reported previously [29,31]. These include good feasibility (high completion rate with little missing data), internal consistency of subscales (Cronbach’s alphas > 0.75 and confirmatory factor analysis), test-retest reliability (intraclass correlation coefficients: median 0.85, lower quartile 0.81, upper quartile 0.90), convergent validity (substantial correlations with related instruments: GLQ-8, GLQ-Uniscale [33], Priestman and Baum LASA scales [34], and Life Satisfaction Index-A [39]), discriminative ability (between groups with different disease severity), and responsiveness to change within individuals.

Statistical Methods

The scoring algorithm was produced by modeling the valuation survey data using the multistep approach developed by Lumley et al. [40]. A “subset index” is calculated by combining the questionnaire subscales into a subset index according to weights based on their correlations with a global scale. Here we define a global scale as a single item asking respondents directly for a unified assessment of their HRQL [41]. An “overall HRQL index” is then calculated by combining the subset index with this global scale using weights based on their statistical precision. Finally, a “utility index” is calculated by transforming the overall HRQL index. A novel feature of Lumley’s approach that is not incorporated in other approaches to deriving scoring algorithms [1,2,42] is that it combines a single-item global scale with multi-item subscales for specific aspects of HRQL. The purpose of including the global scale in the index is to incorporate information about any additional aspects of HRQL that are important but not captured by the subset index [40].

The scoring algorithm for the UBQ-C was derived in four steps (Fig. 1). First, subscale scores for physical function, social/usual activities, self-care, and distresses were calculated from the ratings on the relevant UBQ-C items. Second, a subset index was calculated by weighted combination of the subset scores. Third, an overall HRQL index was calculated by weighted combination of the subset index and the health status thermometer. Fourth, the overall HRQL index was converted to a utility-based index with a suitable transformation. The following paragraphs describe each step in detail.

The subset scores for physical function, social/usual activities, self-care, and distresses are the simple averages of the relevant, nonmissing items, linearly transformed to a scale from 0 (worst) to 1 (best). Responses to the items about “Sex life” and “Other problems” are not included when calculating the scores for the subscales because they are commonly omitted by respondents.

The subset index was calculated by weighted combination of the subscales for physical function, social/usual activities, self-care, and distresses. Weights for the subscales (W1–4 in Fig. 1) were derived from, and proportional to, the coefficients obtained from multivariable ordinary least squares linear regression of the health status thermometer on the subscales. The weights are designed to reflect the relative contribution of each subscale to overall HRQL. The scores for the subset index for each subject were calculated by applying the weights to the subscale scores as follows:

\[ \text{Subset index} = \frac{W1 \times \text{PF}}{W1} + \frac{W2 \times \text{SA}}{W2} + \frac{W3 \times \text{SC}}{W3} + \frac{W4 \times \text{DI}}{W4} \]

(W1–4 are the weights for the subscales: PF is physical function, SA is social/usual activities, SC is self-care, DI is distresses. The score for the subset index was recorded as missing if any of its component scores were missing.

The overall HRQL index was calculated by weighted combination of the subset index with the health status thermometer.
Greater weight was given to the component with least measurement error. The weights were calculated using Lumley’s formula, as follows:

\[ W = \frac{\text{Var}(T) \times (1 - r(T))}{\text{MSE}(R)} \]  

(2)

W is the weight allocated to the subset index, so \( 1 - W \) is the weight allocated to the health status thermometer (Fig. 1). \( \text{Var}(T) \) is the variance of the health status thermometer obtained from the dataset. \( r(T) \) is the intraclass correlation coefficient of the health status thermometer, and was calculated with test–retest data from a previous validation study [29]. \( \text{MSE}(R) \) is the mean square for error from the linear regression of the health status thermometer on the four subscales, and was obtained from the dataset. The scores for the overall HRQL index for each subject were calculated as follows:

\[ \text{Overall HRQL index} = [W \times \text{Subset index}] + [(1 - W) \times \text{HST}] \]  

(3)

HST is the health status thermometer. The scores for the overall HRQL index were recorded as missing if the score for the subset index or health status thermometer was missing.

A suitable transformation function was sought to convert the overall HRQL index to the utility index. We considered a range of functional forms used to transform measures of HRQL to measures of utility in previous studies [43]. We selected the function that best mapped the relationship between the overall HRQL index and TTO utility in the development dataset. The scores for the utility index for each subject were calculated by applying the chosen transformation function to the scores for the overall HRQL index.

Preliminary validation of the algorithm was performed by comparing the scores on the utility index to those from other measures of HRQL, health status, and utility. We assessed how closely the utility index was related to the TTO utility using Spearman’s rank-order correlation (\( r_s \)) and paired \( t \) tests. Associations between the utility index and two independent global measures of HRQL, the general health item from the SF-36 (referred to above) and the Spitzer-Uniscale of global life quality [44, 45], were also assessed with Spearman’s rank-order correlation. We tested the hypothesis that compared with related global measures, the derived indices would give estimates of differences in mean scores between subjects that were grouped by their response to the general health item that were more precise (narrower confidence intervals) but unbiased (similar point estimate).

The overall HRQL index was compared with the health status thermometer, and the utility index was compared with the TTO utility. Differences in mean scores between groups were calculated using unpaired \( t \) tests. The relative precisions of the related measures were compared using the relative efficiency statistic [46, 47]. 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 from the comparison of the groups using the derived index divided by the \( t \)-score from the comparison of the groups using the related global measure.

Results

The study profile describing the subjects and data used to generate the scoring algorithm is shown in Figure 2. Of the 323 patients that were approached to take part in the study, 204 were eligible. Compliance was excellent, with planned interviews and questionnaires completed by 98% of participants. All items on the UBQ-C, except “Sex life” and “Other problems,” were completed by over 90% of patients. Characteristics of eligible sub-
Subjects are summarized in Table 1. All patients had advanced cancer, mostly arising from the breast or bowel. The mean age was 56 and most age groups were represented. The different levels of general health status were also well represented. Most modes of treatment were represented: chemotherapy, supportive care, and observation.

UBQ-C ratings on the health status thermometer and subscales are summarized in Table 2. Patients reported worst impairment for physical function and least impairment for self-care.

The derived weights for the subscales (W1–4), health status thermometer (W), and subset index (1-W) are shown in Table 3. The health status thermometer accounted for about two-thirds of the index for overall HRQL. Of the subscales, greatest weight was given to distresses and least to self-care.

The overall HRQL index was calculated by applying these weights to the subjects’ ratings on the UBQ-C using formulae (1) and (3). The transformation that best reflected the relationship between the overall HRQL index and TTO utility was a disutility power transformation, viz:

\[
\text{Utility index} = 1 - (1 - \text{overall HRQL index})^{2.01} \tag{4}
\]

This transformation was used to convert the overall HRQL index into the utility index.

Scores for the overall HRQL index, utility index, and TTO utility are compared in Table 4. The TTO utility was 1.0 for about half the subjects, despite significant impairments in HRQL. Because of this skewed distribution, the mean value of the TTO utility was lower than its median value. There were no

<table>
<thead>
<tr>
<th>Table 1 Patient characteristics (N = 204)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cancer type</strong></td>
</tr>
<tr>
<td>Breast</td>
</tr>
<tr>
<td>Bowel</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
</tr>
<tr>
<td>&lt;40</td>
</tr>
<tr>
<td>40–49</td>
</tr>
<tr>
<td>50–59</td>
</tr>
<tr>
<td>60–69</td>
</tr>
<tr>
<td>≥70</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
</tr>
<tr>
<td>Partner</td>
</tr>
<tr>
<td>No partner (single, divorced, widowed)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
</tr>
<tr>
<td>Primary school</td>
</tr>
<tr>
<td>Some high school</td>
</tr>
<tr>
<td>Completed high school</td>
</tr>
<tr>
<td>Higher education</td>
</tr>
<tr>
<td><strong>Country of origin</strong></td>
</tr>
<tr>
<td>Australia</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td><strong>General health</strong></td>
</tr>
<tr>
<td>Excellent</td>
</tr>
<tr>
<td>Good</td>
</tr>
<tr>
<td>Fair</td>
</tr>
<tr>
<td>Poor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2 Ratings on the Utility-Based Questionnaire-Cancer health status thermometer and subscales</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>Health status thermometer</td>
</tr>
<tr>
<td>Physical function</td>
</tr>
<tr>
<td>Social/usual activities</td>
</tr>
<tr>
<td>Self-care</td>
</tr>
<tr>
<td>Distresses</td>
</tr>
</tbody>
</table>

All ratings on scale from best (one) to worst (zero).

<table>
<thead>
<tr>
<th>Table 3 Weights for the health status thermometer, subset index, and subscales</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weight</strong></td>
</tr>
<tr>
<td>W</td>
</tr>
<tr>
<td>1-W</td>
</tr>
<tr>
<td>W1</td>
</tr>
<tr>
<td>W2</td>
</tr>
<tr>
<td>W3</td>
</tr>
<tr>
<td>W4</td>
</tr>
</tbody>
</table>

Figure 2 Study profile.
Table 4  Comparison of scores for overall health-related quality of life (HRQL) index, utility index, and time trade-off utility

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Time trade-off utility</th>
<th>Overall HRQL index</th>
<th>Utility index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.91</td>
<td>0.74</td>
<td>0.92</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.17</td>
<td>0.14</td>
<td>0.08</td>
</tr>
<tr>
<td>95% confidence intervals</td>
<td>(0.89, 0.94)</td>
<td>(0.72, 0.76)</td>
<td>(0.90, 0.93)</td>
</tr>
<tr>
<td>Median</td>
<td>1.00</td>
<td>0.77</td>
<td>0.95</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>(0.88, 1.00)</td>
<td>(0.63, 0.86)</td>
<td>(0.87, 0.98)</td>
</tr>
<tr>
<td>% with score of 1.0</td>
<td>50</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Mean difference compared</td>
<td>Not applicable</td>
<td>0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>to time trade-off utility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P value</td>
<td>&lt;0.0001</td>
<td></td>
<td>0.6</td>
</tr>
</tbody>
</table>

All ratings on scale from worst (0) to best (1).

associations between the TTO utility and the patient characteristics listed in Table 1 (data not shown).

The overall HRQL index gave substantially lower scores than the TTO utility (means 0.74 vs. 0.92, difference 0.17, 95% confidence interval [CI] 0.14–0.19). Scores were similar for the utility index and the TTO utility (means 0.92 vs. 0.91, difference 0.01, 95% CI –0.02–0.03).

Comparisons of the utility index with other measures of HRQL, utility, and health status provide preliminary evidence of its validity. The utility index was moderately correlated with the TTO utility (r = 0.38), the general health status item from the SF-36 (r = 0.63), and the Spitzer-Uniscale of global life quality (r = 0.68). The estimated differences in mean scores between subjects grouped by general health in the development dataset were more precisely estimated by the derived indices than by the health status thermometer or the TTO utility (Fig. 3). The relative efficiency statistics in Figure 3 correspond with reductions in sample size needed to detect a significant difference by using the indices of 33% for the overall HRQL index compared with the health status thermometer, and of 75% for the utility index compared with the TTO utility.

Conclusions

We have derived a scoring algorithm for a disease-specific utility-based instrument that is designed to inform clinical decisions about cancer treatments. The algorithm converts ratings from a cancer-specific HRQL questionnaire called the UBQ-C into a utility-based index. Firstly, the algorithm calculates a subset index from a weighted combination of the UBQ-C subscales for physical function, social/usual activities, self-care, and distresses. Secondly, an overall HRQL index is calculated from a weighted combination of the health status thermometer and the subset index. Thirdly, the algorithm calculates a utility index by applying a power transformation to the overall HRQL index. The scoring algorithm was developed using TTO utilities and UBQ-C ratings elicited from patients with advanced cancer who rated their current health status and HRQL. The utilities can be used to generate QALYs to compare cancer treatments.

Utilities and QALYs are a useful way to compare cancer treatments because they can be evaluated on a common scale that incorporates disparate treatment effects like gains in survival duration, improvements in HRQL because of relief of cancer symptoms, and deteriorations in HRQL as a result of treatment-related side effects [1,48]. Analyses of cancer trials in terms of utilities and QALYs are increasingly used to inform economic decisions about cancer treatments [49–56], but can also be used to inform clinical decisions [57–62]. Despite the advantages of utilities and QALYs, there is no standardized approach for eliciting utilities [1,24,63,64]. One way to obtain utilities is to elicit them directly from respondents using a standard gamble or TTO interview, but this task is complex, resource intensive, and can be distressing if cancer patients are required to assign utilities for their own health state [24]. A more practical approach is to derive utility scores from a utility-based instrument. We posit that deriving utility scores from a utility-based instrument that is disease-specific and based on the perspective of patients is the best approach for informing clinical decisions.

The UBQ-C is a disease-specific instrument that is designed for the evaluation of cancer treatments. It asks about important consequences of cancer and its treatment not covered by generic instruments such as the EQ-5D [3], HUI3 [7], or SF-6D [8] including fatigue, nausea, shortness of breath, and hair loss. The main advantage of a disease-specific instrument such as the UBQ-C over a generic instrument for generating utility scores is that it should be more sensitive to detect differences in HRQL between individuals with cancer. This requires empirical testing,
as has been performed for other disease-specific instruments that
generate utility scores for cancer [12,15,65] and a range of other
diseases including bladder disorders [6,16], hearing impairment
[66], and asthma [13]. Another advantage of using a disease-
specific utility-based instrument is that it provides data on specific
aspects of HRQL, overall HRQL and utility with a single ques-
tionnaire, and increases the availability of utility data for com-
parisons of treatment from randomized clinical trials [15]. This
approach enables utilities to be derived from previous studies
where the UBQ-C was used, and reduces questionnaire burden for
future trial participants by having a single questionnaire and
approach that provides these three kinds of information.

The major limitation of using disease-specific, utility-based
instruments is that the utility scores they provide may not be
comparable to those derived from other instruments, particularly
generic instruments, because the dimensions of health status and
HRQL that they cover are different [9–11]. Whether this is a
problem depends on the decision for which the utilities are being
applied. We argue that disease-specific instruments are best suited
to treatment comparisons within a particular disease used to
inform clinical decisions. In this context, comparisons across
other diseases and healthy populations are less important, but
coverage of aspects relevant to the patients under study is crucial.
Others have argued that disease-specific instruments may also be
suitable for treatment comparisons across all diseases to inform
health funding and policy decisions if the scoring algorithm is
derived using a valuation technique and population sample that
is similar to a generic instrument, and the utility scores are shown
to be comparable [6].

The algorithm described in this study was based on the per-
spective of cancer patients who were currently experiencing those
health states. The perspective differs in two important ways from
scoring algorithms used for most of the generic and cancer-
specific utility-based instruments reported previously. First, it is
the perspective of patients rather than lay people. Second, it
reflects views about a health state that is real and current rather
than hypothetical and in the future [67]. The perspective from
which a utility is elicited may have significant implications for
clinical and economic decisions that incorporate utilities and
QALYs, because patients typically assign higher utilities to a
given health state than lay people [18–20]. This may reflect partly
the lay person’s difficulty appreciating what a hypothetical health
state is really like, and partly the patient’s adaptation to their
own health state [18,24,25,60,68]. The choice of perspective
should reflect the viewpoint from which the results will be inter-
preted [24,25]. As discussed in the introduction, it is generally
agreed that the perspective of patients is more appropriate for
informing clinical decisions about specific treatments, while the
perspective of lay people is more appropriate for informing deci-
sions about health policy and funding. Some also argue that the
perspective of patients should be used to inform funding and
policy decisions, because patients better understand what it is
like to live with a particular disease [17,19], but this argument is
controversial because it runs counter to prevailing health eco-
omic guidelines [21,23,69].

This study also provides preliminary evidence supporting the
validity of the utility index. It was substantially correlated with
independent measures of general health, overall life quality, and
TTO utilities. Mean scores for groups from the utility index and
TTO utility were almost identical. This supports the validity of
using the utility index to generate mean utilities for comparing
patient groups. However, as expected, we found that the utility
index did not accurately predict utilities for individual patients.
The mean absolute difference between the utility index and TTO
utility for each subject was relatively large at 0.10. This finding
argues against using the utility index to predict utilities for indi-
viduals. This is exactly as expected [15], because utilities are
influenced by factors apart from HRQL such as individuals’
attitudes to risk and uncertainty [24,64].

The derived indices for HRQL and utility gave more precise
estimates of differences between groups than the health status
thermometer or TTO utility. We expected more precise estimates
because any score aggregated from multiple items will produce a
more precise estimate of differences between groups than a
single-item scale [40,70]. This finding does not strengthen or
weaken the validity of the indices but is an expected measure-
ment property which enhances the sensitivity and responsiveness
of the indices.

Ongoing work is needed to support the validity of the utility
index. A companion article describes the application of the
scoring algorithm to independent trial datasets in breast cancer,
and provides further evidence to support its validity by compari-
son with clinical data [28]. We have also reported on a compari-
son of the value and sensitivity of utility scores generated by
the indices to those generated by the EQ-5D in colorectal cancer
[71]. Independent testing in other datasets will further establish
validity.

The study population and valuation survey used to develop
the scoring algorithm has several strengths. The patient charac-
teristics were diverse including men and women with a broad
range of ages, levels of performance status, and levels of health
status. Compliance was excellent with both UBQ-C completion
and utility interviews. We used the TTO method to elicit utilities
for health states. The TTO is practical, reliable, and has empiri-
cal validity [24,72,73]. A limitation of our valuation survey was
that its sample size was too small to allow division of the group
into a “training” set, where the algorithm was developed, and a
“validation” dataset where its validity and accuracy was inde-
pendently tested. The dataset was confined to patients with
advanced cancer, mostly with breast or colorectal primaries, and
attending ambulatory clinics. This may raise questions about the
generalizability of the algorithm and approach to patients with
cancers that are of earlier stage, in remission, or from other
primary sites.

The novelty of our statistical approach is in its combination
of a single-item global scale with multi-item subscales for specific
aspects of HRQL and its methods for deriving optimal weights.
Most other utility-based indices do not incorporate a single-item
global scale [1,2,42]. Incorporation of the single-item global
scale has two potential advantages. First, it provides a unified
reflection of how the patient rates their health status that enables
inclusion of aspects of HRQL that are important but are not
directly captured by multi-item subscales [40]. Second, it allows
the scoring algorithm to be optimized in different treatment
contexts by adjusting the weights assigned to the multi-item
subscales [40]. The purpose of optimizing the algorithm is to
reflect the differences in importance that patients with different
types and stages of cancer, and treatments assign to various
dimensions of HRQL [40,74]. The implications of optimizing the
algorithm for different treatment contexts are addressed by appli-
cation and discussion in a companion article [28].

This work enables HRQL data obtained with a simple cancer-
specific questionnaire to be converted into a utility index that
reflects the perspective of cancer patients. The approach is best
suited to generating estimates of mean utilities for groups, and
our work so far supports this application. It can be applied in
clinical trials to compare the effect of cancer treatments on
HRQL using utility measures, and to generate QALYs for
informing clinical decisions and as an alternate viewpoint for
economic analyses. The approach provides a general method
for converting HRQL ratings to valid utility-based measures that could be applied in other trial settings for analysis of HRQL data collected with different questionnaires.

Source of financial support: Peter Grimison was supported by National Health and Medical Research Council, Australia; Cancer Institute NSW, Australia; and GlaxoSmithKline 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 article, or the decision to submit the article for publication.

Supporting information for this article can be found at: http://www.ispor.org/publications/value/ViHsupplementary.asp

References

10 Dowie J. Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. Health Econ 2002;11:1–8.
11 Feeny D. Commentary on Jack Dowie, “Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions.” Health Econ 2002;11:13–6.