Article:
King, MT, Costa, DSJ, Aaronson, NK et al. (12 more authors) (2016) QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. Quality of Life Research, 25 (3). pp. 625-636. ISSN 0962-9343

https://doi.org/10.1007/s11136-015-1217-y

QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30

King MT[1,2], Costa DSJ[1], Aaronson NK[3], Brazier JE[4], Cella DF[5], Fayers PM[6], Grimison P[7], Janda M[8], Kemmler G[9], Norman R[10,11], Pickard AS[12], Rowen D[4], Velikova G[13], Young TA[4] , Viney R[11], on behalf of the MAUCa Consortium.

[1]Psycho-Oncology Cooperative Research Group (PoCoG), University of Sydney, Sydney, Australia.

[2]Central Clinical School, Sydney Medical School, University of Sydney, Sydney, Australia.

[3]The Netherlands Cancer Institute, Amsterdam, The Netherlands.

[4]School of Health and Related Research, University of Sheffield, Sheffield, South Yorkshire, UK.

[5]Northwestern University, Chicago, IL, USA.

[6]University of Aberdeen, Aberdeen, UK, and Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

[7]Chris O Brien Lifehouse Sydney Medical School University of Sydney, Sydney, Australia.

[8]School of Public Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, QLD, Australia.

[9] Innsbruck Medical University, Innsbruck, Austria

[10]School of Public Health, Curtin University, Perth, Australia

[11] Centre for Health Economics Research & Evaluation (CHERE), University of Technology Sydney (UTS), Sydney, Australia.

[12] Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA.

[13]St James's Hospital, Leeds, UK.

Target Journal: Quality of Life Research

Corresponding author:
Prof Madeleine King
madeleine.king@sydney.edu.au
Tel +61 2 9036 6114
Fax +61 2 9036 5292

Keywords: Quality of life; utility; QLQ-C30; multi-attribute utility instrument; cancer

Acknowledgements The MAUCa Consortium, in addition to those named as authors, consists of the following members, all of whom made some contribution to the research reported in this paper: Stuart Peacock, Helen McTaggart-Cowan, Julie Pallant and Deborah Street. We would also like to thank the following people for their generosity in contributing data for secondary
analysis: U Abacioğlu J Arraras J Blazeby W-C. Chie, S. Clarke, S. Kaasa, P. Klepstad, Millennium Pharmaceuticals, K. Mystakidou, S. Peacock, R. Schwarz, N. Scott, N. Tebbutt, G. Velikova and the Australian Gastro-Intestinal Trials Group. This research was supported by a National Health and Medical Research Council (Australia) Project Grant (632662). A/ Professor Janda was supported by a NHMRC career development award 1045247.
Professor King was supported by the Australian Government through Cancer Australia.

ABSTRACT

Purpose: To derive a health state classification system (HSCS) from the cancer-specific quality of life questionnaire, the EORTC QLQ-C30, as the basis for a multi-attribute utility instrument.

Methods: The conceptual model for the HSCS was based on the established domain structure of the QLQ-C30. Several criteria were considered to select a subset of dimensions and items for the HSCS. Expert opinion and patient input informed *a priori* selection of key dimensions. Psychometric criteria were assessed via secondary analysis of a pooled dataset comprising HRQOL and clinical data from 2616 patients from eight countries and a range of primary cancer sites, disease stages, and treatments. We used confirmatory factor analysis (CFA) to assess the conceptual model's robustness and generalisability. We assessed item floor effects (>75% observations at lowest score), disordered item response thresholds, coverage of the latent variable and differential item function (DIF) using Rasch analysis. We calculated effect sizes for known group comparisons based on disease stage and responsiveness to change. Seventy-nine cancer patients assessed the relative importance of items within dimensions.

Results: CFA supported the conceptual model and its generalizability across primary cancer sites. After considering all criteria, 12 items were selected representing 10 dimensions: physical functioning (mobility), role functioning, social functioning, emotional functioning, pain, fatigue, sleep, appetite, nausea, bowel problems.

Conclusions: The HSCS created from QLQ-C30 items is known as the EORTC Quality of Life Utility Measure-Core 10 dimensions (QLU-C10D). The next phase of the QLU-C10D's development involves valuation studies, currently planned or being conducted across the globe.

INTRODUCTION

The European Organisation for Research and Treatment of Cancer (EORTC) has a modular approach to the assessment of health-related quality of life (HRQOL) [1]. Their core questionnaire, the QLQ-C30 [2], is one of the most widely used patient-reported outcome measures in cancer clinical trials. The scoring algorithm produces 15 scales [3]: five key aspects of functioning; a range of symptoms commonly experienced by cancer patients; and a global assessment of HRQOL. In this way, the QLQ-C30 provides a comprehensive profile of outcomes that are important to patients and their health care providers. However, because the QLQ-C30 is a HRQOL profile measure, not a preference-based measure, it cannot be used in cost-utility analysis (CUA) [4].

CUA is now required or preferred in many jurisdictions for health technology assessment and health reimbursement decisions [5-8]. CUA is a cost-effectiveness analysis that uses a health outcome metric called utility to weight survival, typically as the Quality Adjusted Life Year (QALY). The utility metric has three key features that distinguish it from the scales of the QLQ-C30 and other HRQOL profile measures. First, key dimensions of HRQOL are integrated into a single index. Second, it is interpreted as a cardinal scale with two key anchor points: 1 is equivalent to full health (the maximum possible value) and 0 is equivalent to being dead (health states worse than death have negative values) [4]. Third, obtaining utilities requires stated preference-based assessment methods, or valuation tasks [9], i.e. standard gamble [10], time trade-off [11] or discrete choice experiments [12; 13]. These tasks typically pose implicit or explicit trade-offs between dimensions of HRQOL and survival, which are the empirical basis of the weighting of HRQOL dimensions in the utility index.

Some preference-based measures have been derived from HRQOL profile measures [14-17], typically in two stages. First, a subset of dimensions and items is selected from the HRQOL measure to form a health state classification system (HSCS). This is required because non-preference-based HRQOL measures typically include more items than is manageable in the valuation task required for the second stage, in which a sample of health states is valued and an algorithm derived for estimating the utility of all possible health states. The outputs of these two stages - a HSCS and a utility scoring algorithm - constitute a multi-attribute utility instrument (MAUI). This paper describes the first step in developing an internationally valid, cancer-specific MAUI based on the QLQ-C30. The second step is addressed in a companion paper[18].

METHODS

To derive a HSCS from the QLQ-C30, we adapted and extended methods used previously [17; 19]. This involved determining the core dimensions, then applying nine criteria to select items within dimensions. The rationale and methods for each criterion are described below. The research was conceived and conducted by the Multi-Attribute Utility in Cancer (MAUCa) Consortium, and coordinated from the University of Sydney (Human Research Ethics Committee approval 13207).

The EORTC QLQ-C30 (version 3)

This is a multidimensional questionnaire composed of 30 items which form five functioning scales, three multi-item symptom scales, five single-item symptom scales plus a financial difficulties item and a two-item global HRQOL scale [3]. Its development was informed by extensive literature review, stakeholder input, psychometric evaluation and field testing [1; 2], and its validity and reliability are well established [2; 20]. The 28 functioning and symptom items are rated on a four-point scale (1 = "Not at all", 2 = "A little", 3 = "Quite a bit", 4 = "Very much").

Core dimensions of the HSCS – *a priori* inclusions and exclusions

The authors met in October 2010 to determine the core dimensions of the HSCS. We represent a range of relevant expertise: health economics (JEB, RN, SP, ASP, DR, RV, TAY), HRQOL in cancer (NKA, DFC, PMF, MTK, GV), psychometrics (DSJC, PMF, JFP), behavioural science (MJ, NKA) and oncology (PG, GV). Given the robustness of the QLQ-C30's development process, we agreed that its established domain structure provides a good basis for the conceptual model for the HSCS, but that 15 dimensions (i.e., QLQ-C30's 15 scales) would be unmanageable in subsequent valuation tasks. The two global health/QOL items were excluded because conventionally the attributes in MAUIs are specific domains of health. The Financial Concerns item was excluded as it describes a consequence rather than an aspect of health. Four functioning domains were considered essential – physical, role, emotional and social. Cognitive functioning was deemed non-essential for CUA in cancer, and is the least reliable scale [2], so was recommended for exclusion, pending results from other criteria. We agreed that six common symptoms of cancer (pain, fatigue, nausea/vomiting, gastrointestinal disturbance, appetite loss, sleep disturbance) were likely to be important in CUA in cancer, and may provide greater sensitivity (to differences between groups) and responsiveness (to change over time) than generic MAUIs. The remaining symptoms were deemed of lower priority, with decisions about inclusion/exclusion pending results from other criteria. Following *dimension* selection, we used the quantitative methods described below to guide *item* selection.

Data

Criteria 1-8 (below) were evaluated via secondary analysis of datasets containing QLQ-C30v3, and patient age, sex, primary cancer site, stage and treatment. Criteria 1-7 required only one observation per patient, Criteria 8 required two observations per patient. Suitable datasets were sought via MAUCa Consortium members and their associates.

Categorical variables were coded as follows:

- Primary cancer site (15 categories): breast; colorectal; genitourinary; gynaecological; head and neck; leukaemia; hepato-biliary; lung; melanoma; myeloma; gastro-oesophageal; prostate; sarcoma; testicular; other.
- Disease stage (2 categories): loco-regional; recurrent/metastatic.

- Timing of assessment (3 categories): either baseline (prior to the start of a course of treatment), on treatment (during a course of treatment), or follow-up (during clinical follow-up after a course of treatment).
- Treatment (17 categories): no treatment; chemotherapy only; radiotherapy only; hormone therapy only; surgery only; analgesics only; then a further 11 categories for various combinations of those treatments.

Criterion 9 (patient opinion) required primary data collected using a study-specific self-report survey. Cancer patients were recruited from four Australian hospitals: three in regional New South Wales and one in metropolitan Queensland.

Data analysis

Confirmatory Factor Analysis (CFA)
CFA was used to confirm the QLQ-C30 measurement model prior to selecting items from each factor for the HSCS; technical details are reported elsewhere[21]. Given the QLQ-C30 items are ordinal, the mean- and variance-adjusted weighted least squares (WLSMV) estimation method was used, implemented with MPlus. Standardised factor loadings were examined. Measurement invariance of the factor structure across primary cancer sites was tested with multi-group CFA, as reported elsewhere[22].

The measurement model
A measurement model consists of manifest (observed) variables, postulated to be indicators of one or more latent (unobserved) variables. At least two items are needed to estimate each latent variable, so only the multiple-item dimensions were included: Physical Functioning (PF, items 1-5); Role Functioning (RF, items 6-7); Emotional Functioning (EF, items 21-24); Social Functioning (SF, items 26-27); Cognitive functioning (CF, items 20 and 25); Pain (items 9 and 19); Fatigue (items 10, 12 and 18); Nausea and Vomiting (items 14 and 15).

Criteria to assess items within dimensions
The aims of Criteria 1-8 were to identify problems with items that might justify exclusion from the HSCS. Criteria 1-5 were assessed with Rasch analysis for each dimension confirmed with CFA, implemented in RUMM2030 [17] [23]. The criteria are summarised below.

1. Fit of items to the Rasch model: Overall fit was assessed by examining the chi-squared statistic with a Bonferroni correction. Misfit due to persons or items (>1.5 the standard deviation of the fit residual) was further assessed via fit residuals for individual items or persons (items with fit residuals >2.5 were removed; persons with fit residuals >2.5 were removed only if they appeared to contribute to item misfit). This process was repeated until only well-fitting items remained and the overall goodness of fit statistic was non-significant.

2. Disordered response thresholds: An appropriately functioning item requires a response format that respondents use in a consistent manner. We assessed this by examining item-threshold probability curves.

3. Spread of item thresholds across the latent variable: An item with thresholds that cover a wide range of the latent variable is a better single representative of the dimension than an item with thresholds within a narrow range. We inspected the item maps, and present each item's lowest and highest response threshold as summary statistics.

4. Differential item function (DIF): a form of bias in which systematic differences in patterns of responding to an item are observed between individuals with different characteristics, despite having the same level of the latent variable. We examined DIF by sex and cancer site.

5. Local dependence: For any pair of items with residual correlation ≥0.3 above the mean residual correlation for all pairs, we considered forming a composite item for inclusion in the HSCS.

6. Floor and ceiling effects: Frequencies of response categories for each item were examined for floor or ceiling effects. Items exhibiting either effect were considered poor candidates for the HSCS, either because they are uncommon and therefore unlikely to affect patients' HRQOL, or because they may be unresponsive to treatments.

7. Sensitivity to differences between early and late stage cancer: For each item, Cohen's measure of effect size, *d*, was calculated as the mean of late stage patients minus the mean of early stage patients divided by the pooled standard deviation for these two groups.

8. Responsiveness of items to change due to treatment: Assessing responsiveness requires two observations per patient over a period when HRQOL is expected to change in a systematic way [24]. As clinical context differed markedly across the available datasets, we used change from baseline to on-treatment as the anchor for expected change, and calculated responsiveness indices for each study separately as the mean difference between baseline and on-treatment observations divided by baseline standard deviation (effect size [25]). Datasets were considered suitable for assessing responsiveness if change of at least 0.2SD was observed in at least one HRQOL domain score.

9. Patient opinion: We sought cancer patients' opinions about the relative importance of items. We developed a survey, refined iteratively through a series of "think aloud" interviews with 9 patients in a large Australian metropolitan hospital, to ensure that self-completing participants understood the cognitive task required[26]. The survey was subsequently administered to 79 patients in four Australian hospitals (one metropolitan, three rural). QLQ-C30 items were presented in five groupings: physical functioning (5 items); emotional functioning (4); role,

cognitive and social functioning (6); fatigue, pain, nausea/vomiting (7); dyspnoea, sleep disturbance, appetite loss, constipation, diarrhoea (5). Respondents were instructed to tick the three items in each group that had affected their HRQoL most since their cancer was diagnosed ('Top 3'), and then select the one that had affected their HRQoL most ('Most Important'). We calculated the frequency with which each item was nominated as 'Most Important' and 'Top 3'.

Final decisions after considering all criteria
The authors met in October 2012 to consider results for Criteria 1-8, and to make penultimate decisions on item inclusion/exclusion. During 2013, the patient survey and DCE methods experiment were conducted and their results considered in finalizing decisions about the content of the HSCS.

RESULTS

Eighteen suitable datasets were obtained [27-44]. Table 1 summarises their characteristics and the number of observations each contributed to various psychometric analyses.

A pooled dataset of 2,616 observations (one per patient) was used for CFA and analyses for Criteria 1-7. All commonly occurring cancers are represented, with 40% with localized/regional and 60% recurrent/metastatic (Table A, online appendix). All common treatments were represented, with 54% of observations related to chemotherapy, 30% to radiotherapy, 27% to surgery, 4% to hormone therapy; 65% related to a single therapy, 26% to multiple therapies, and 9% to no therapies (Table B, online appendix). In 22%, HRQOL was assessed prior to therapy (baseline), 44% during therapy, and 33% during follow-up. For Criterion 8, nine datasets containing a total of 749 patients with both baseline and on-treatment observations, were used to estimate responsiveness (Table C, online appendix). Patient illness and treatment characteristics of patients in the Patient Opinion substudy (used to assess Criterion 9), are presented in Table D (online appendix).

Table 2 summarises results for CFA and all nine criteria. Item maps (Criterion 3) are presented in Figures A-H (online appendix). For Criterion 6, Figure I (online) shows that only items 5 and 15 displayed floor effects, with approximately 80% of respondents selecting "Not at all". Detailed results for Criterion 8 (responsiveness) are available in Table D (online). We describe below how these results informed decisions about which items to retain in the HSCS. Table 3 presents the dimensions and levels in the resultant HSCS.

Confirmatory Factor Analysis
The loadings of all items on their respective factors were relatively strong and statistically significant ($p<0.001$). Model fit was adequate (CFI=0.98, TLI=0.98, RMSEA=0.071).

Rasch Analysis

Within their dimensions, no item exhibited poor fit to the Rasch model or local dependence. One item exhibited a disordered item response threshold and two items exhibited DIF. Thus Criteria 1, 2, 4 and 5 generally played little role in identifying items for exclusion from the HSCS.

Physical Functioning
Item 5's disordered response threshold was remedied by combining the middle two response categories, after which model (p=0.015) and person (0.71) fit were good, and item fit was marginal (1.53). The item map (Figure A) reflected the Guttman nature of this set of items, with Item 1 thresholds clustered at the good functioning end of the latent variable and Item 5 thresholds clustered at the poor end. Taken together, the two mobility items (Items 2-3) provided good coverage of the latent construct, as reflected in the range from the lowest response threshold for Item 2 to highest threshold for Item 3. Both items had good sensitivity and responsiveness, and together accounted for the majority of patients' ratings for the 'Top 3' and 'Most Important' items. Items 2 and 3 were therefore chosen as a pair to represent this dimension, with 4 levels defined as in Table 3.

Role functioning
Rasch model (p=0.612), item (0.09) and person (1.36) fit were all good. There was little to distinguish the two role functioning items in terms of psychometric criteria: both gave similar coverage of the latent variable, both exhibited DIF by primary cancer site, and both had similarly good sensitivity and responsiveness. As Item 6 was commonly chosen as 'Most Important' or 'Top 3' by patients, it was selected to represent this dimension in the HSCS.

Emotional functioning
Rasch model, item and person fit were all good (0.894, 0.86 and 1.19, respectively). These four items provided similar coverage of the latent variable. Although none were particularly sensitive, all but Item 22 had generally good responsiveness. Item 22 (worry) was selected to represent this dimension in the initial valuation task [18], largely because it was most commonly chosen as 'Most Important' or 'Top 3' by patients. However, after reviewing the results from the valuation task [18], this decision was changed to Item 24 (depressed), for reasons outlined in the discussion.

Social functioning
Rasch model, item and person fit were all good (0.503, 0.77 and 0.62, respectively). Item 26 (family life) provided a wider range than Item 27 (social activities) and was chosen more often as 'Most Important' or 'Top 3' by patients. To avoid disadvantaging people without family, we combined both items in the HSCS.

Cognitive functioning
Rasch item (0.51) and person (1.00) fit were good, but model fit (p=0.0006) was poor, possibly due to the different threshold locations of the two items, with Item 25 providing the greatest range. Neither item was selected often as 'Most Important' by patients when grouped with role and social functioning, confirming the team's initial inclination to exclude this dimension from the HSCS.

Symptoms

Rasch model, item and person fit were all good for the Pain (0.026, 0.57 and 1.35, respectively), Fatigue (0.378, 0.36, 1.12) and Nausea/Vomiting (0.099, 0.49, 0.43) dimensions. For fatigue, psychometric properties provided mixed results, with Item 12 providing the best coverage of the latent variable, Item 10 being most sensitive and Item 18 most responsive. As Item 18 was most often chosen as 'Most Important' and 'Top 3' by patients, it was selected to represent this dimension in the HSCS. Item 9 was selected to represent Pain, as Item 19 would have presented a logical inconsistency with Role Functioning Item 6 in certain health states (e.g. "I am not limited in doing work/other daily activities" in combination with "Pain interferes with my daily activities"). For Nausea/Vomiting, Item 14 was superior to Item 15 across all criteria, so was selected.

Among the remaining single-symptom scales of the QLQ-C30, Item 16 (constipation) and 17 (diarrhea) were selected as a pair to represent bowel problems. Both had good responsiveness, and together were chosen quite commonly as 'Most Important' and 'Top 3' by patients. Further, our team of oncologists noted their importance as common symptoms of both cancer and its treatments. Item 11 (trouble sleeping) was very commonly chosen as 'Most Important' and 'Top 3' by patients, so was included. Item 13 (lacked appetite) had good sensitivity and responsiveness, and was considered clinically important by our oncologists.

DISCUSSION

We have conducted extensive analyses using 18 datasets representing 14 countries and all common cancer sites, stages and treatment, to derive a HSCS from the EORTC QLQ-C30. This HSCS is the descriptive part of a cancer-specific MAUI called the QLU-C10D. This name was endorsed by the EORTC QOL Group Executive Committee in April 2014: 'QLU' indicates it is a utility measure; 'C' indicates its origin in the EORTC's core questionnaire; '10D' indicates 10 domains (mobility, role functioning, social functioning, emotional functioning, pain, fatigue, sleep, appetite, nausea, bowel problems). Note that the QLU-C10D is not a brief-form profile instrument, nor a stand-alone measure, but a utility scoring algorithm for use in trials that use the QLQ-C30. The use of a large international pooled dataset representing a wide range of cancers and treatments, and the inclusion of both oncologist and patient opinion, supports the international applicability and clinical validity of the HSCS. However, it contains more dimensions than previous MAUIs except the 15D [45]. To complete the QLU-C10D's development, we have developed a valid and feasible valuation method to generate a preference-based utility scoring algorithm[18]. Country-specific valuation studies are underway in Australia and Europe, and studies in Canada and USA may follow.

Our work extends that of Rowen *et al*, [46], who applied the methods of Young et al [17] to data from 655 multiple myeloma patients to derive an eight-dimensional HSCS. Our HSCS has the same dimensions plus an additional two (Sleep and Appetite). Four of these dimensions are identical (Emotional Functioning, Fatigue,

Nausea, Bowel Problems), one contains the same items but different levels (Physical Functioning) and three contain different items (Role Functioning, Social Functioning, Pain). Pickard et al used similar psychometric item selection methods to derive multi-attribute models in their mapping study [19], obtaining results that had some similarities and some differences from ours and Rowen et al's. These differences in results are likely due to different patient populations, and the item selection process itself, which requires judgment to reach a decision across a range of criteria. Given these differences, we recommend the QLU-C10D be used, particularly as it based on such a broad spectrum of patients and has been endorsed by the EORTC QOL Group.

We adapted and extended the methods of Young et al [17] in two significant ways. First, we used confirmatory factor analysis, which is more appropriate and efficient than exploratory factor analysis in deriving a HSCS from a HRQOL profile measure with a well-established domain structure [47]. Second, we added an innovative aspect to incorporate patients' opinions. This assisted with the decision making where candidate items could not be differentiated using statistical methods. Notably, associated qualitative work revealed that it is often difficult for patients to distinguish severity and importance [48], suggesting that severity is a good proxy for importance in general.

The EORTC QLQ-C30 has been mapped to other preference-based utility measures; the EQ-5D, SF-6D and 15D [49]. As this provides a method to estimate utilities from responses to the QLQ-C30, this raises the question: is there a need for a MAUI developed directly from the QLQ-C30? A directly derived societal value set for the QLQ-C30 is preferable for several reasons. Mapping enables utilities to be predicted from HRQOL profile scores using datasets containing both the profile measure and a validated utility instrument. However, mapping functions assume that the generic MAUI is appropriate for the target cancer population and sensitive to health changes in it; these assumptions are difficult to test empirically and are rarely tested. Thus utilities may be under-predicted if the generic MAUI is insensitive to change, and changes in symptoms measured by the QLQ-C30 may not be reflected in generic utility values. Further, various mapping functions are used, there is no consensus on which is best, utility values vary greatly, and overall model fit is generally only moderate [49]. The mappings are highly dependent on the dataset(s) from which regressions are estimated, and therefore on the coverage of health state space in the patient samples, which are rarely comprehensive. The application of mapping functions which fit the data poorly may result in unrealistic QALY assessments and correspondingly inefficient health resource allocation decisions. Finally, regression generally results in biased estimates [50]; the highest scores are underestimated and the lowest scores are overestimated, with bias increasing as the correlation between the profile measure and the generic MAUI decreases. This was illustrated empirically in a review of 30 mapping studies that showed predicted values from mapping functions had less variance than the original observed values [51]. All these issues support the need for directly elicited preferences for a HSCS based onQLQ-C30, which is the purpose of the QLU-C10D. Decision-making bodies that have considered these issues have placed a lower preference on mapping based approaches [7; 8].

This study has some limitations. All datasets were from western or European countries; further work is required to ascertain applicability in other parts of the world. While most common cancers were included, inevitably some cancers were not, e.g. primary brain cancer. The sample for the patient input component was limited to Australians, with few patients with very advanced disease. Item selection required synthesis across a range of criteria; as there were no strict decision rules about how to weight evidence across criteria, whether it be conflicting or confirming, the final decisions required the team's collective judgment.

Using the QLQ-C30 as a basis for a MAUI HSCS could arguably have some shortcomings, because the QLQ-C30 includes a mixture of proximal (symptoms) and distal (the implications of those symptoms for functioning) domains. If dimensions in a MAUI are strongly causally related, this may compromise assumptions of structural independence, create implausible health states (e.g., high fatigue but high role functioning) and induce double counting. Since the QLU-C10D includes both symptoms and their consequences, the degree of a lack of structural independence for this measure may be more extensive than in other preference-based measures.

The decision to exclude cognitive function from the HSCS was based on clinician and patient input about the relative importance of all dimensions, and our empirical analyses suggested other dimensions were more likely to be sensitive to differences across a wide range of clinical contexts. However, a limitation of our study was that our patient samples did not contain patients with primary brain tumors. Further, patients with serious cognitive impairments are likely to be excluded from quality of life assessment in clinical trials generally. Interestingly, the cognitive functioning scores of our samples were only slightly better than those of the EORTC reference values for brain cancer patients [52]. Although cognitive impairment is emerging as an important consequence of cancer and its treatment, current evidence about "chemo brain" in breast cancer suggests that such impairment tends to be modest [53; 54]. If cognitive functioning is to be given more consideration in health economic decisions, other approaches will be needed.

The development of the QLU-C10D by the MAUCa Consortium represents a significant advance in the incorporation of quality of life information into economic decisions about cancer treatments. Because the QLU-C10D contains symptoms commonly experienced by cancer patients, it may be more sensitive in cancer populations than generic MAUIs, although this will also depend on the values assigned to these symptoms by the general population in valuation studies. Once country-specific value sets are generated, further studies should to examine how the utilities generated by this new cancer-specific MAUI compare to those of generic MAUIs. If the QLU-C10D utility scores are correlated with those of generic MAUIs, and are at least as sensitive and responsive, respondent burden could be reduced by avoiding the need to use a generic utility measure in trials in addition to the QLQ-C30.

Compliance with Ethical Standards

Funding: This research was supported by a National Health and Medical Research Council (Australia) Project Grant (632662). Professor King was supported by the Australian Government through Cancer Australia. Dr Norman was supported by a NHMRC early career research fellowship (1069732).

Conflict of Interest: The authors declare they do not have conflicts of interest.

Ethical Approval: The study was approved by the University of Sydney Human Research Ethics Committee, approval number 2012/2444. All study procedures involving human participants were in accordance with the ethical standards of institutional and national research committees and with the 1964 Helsinki declaration and its later amendments and comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. Aaronson, N. K., Bullinger, M., & Ahmedzai, S. (1988). A modular approach to quality of life assessment in cancer clinical trials. Recent Results in Cancer Research, 111, 231-249.
2. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., deHaes, J. C. J. M., Kaasa, S., Klee, M., Osoba, D., Razavi, D., Rofe, P. B., Schraub, S., Sneeuw, K., Sullivan, M., & Takeda, F. (1993). The European Organisation for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. Journal of the National Cancer Institute, 85(5), 365-376.
3. Fayers, P. M., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., Bottomley, A., & on behalf of the EORTC Quality of Life Group. (2001). The EORTC QLQ-C30 Scoring Manual (3rd ed.). Brussels: European Organisation for Research and Treatment of Cancer.
4. Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. (2005). Methods for the economic evaluation of health care programmes (3rd ed.). Oxford: Oxford University Press.
5. National Institute for Health and Clinical Excellence. (2013). Guide to the Methods of Technology Appraisal. London: NICE.
6. Mittmann, N., W.K., E., Rocchi, A., Longo, C. J., Au, H.-J., Husereau, D., Leighl, N., Isogai, P., Krahn, M., Peacock, S., Marshall, D., Coyle, D., Malfair Taylor, S. C., P., J., & Oh, P. I. (2009). Addendum to CADTH's Guidelines for the Economic Evaluation of Health Technologies: Specific Guidance for Oncology Products. . Ottawa: Canadian Agency for Drugs and Technologies in Health.
7. Mitchell, A. S., & Viney, R. (2010). Meeting the information needs of a national drug payer: Aspirations of the guidelines from Australia. Drug Development Research, vol. 71, no. 8, pp. 463-469., 71(8), 463-469.
8. Scuffham, P. A., Whitty, J. A., Mitchell, A., & Viney, R. (2008). The use of QALY weights for QALY calculations: A review of industry submissions requesting listing on the Australian Pharmaceutical Benefits Scheme 2002-4. PharmacoEconomics, 26(4), 297-310.

9. Neumann, P. J., Goldie, S. J., & Weinstein, M. C. (2000). Preference-based measures in economic evaluation in health care. Annual Review of Public Health, 21, 587-611.
10. Brazier, J., Roberts, J., & Deverill, M. T. (2002). The estimation of a preference-based measure of health from the SF-36. J Health Econ, 21(2), 271-292.
11. Viney, R., Norman, R., King, M. T., Cronin, P., Street, D. J., Knox, S., & Ratcliffe, J. (2014). Time trade-off derived EQ-5D weights for Australia. Value in Health, 14(6), 928-936.
12. Norman, R., Viney, R., Brazier, J., Burgess, L., Cronin, P., King, M., Ratcliffe, J., & Street, D. (2014). Valuing SF-6D health states using a Discrete Choice Experiment. Medical Decision Making, 34(6), 773-786.
13. Viney, R., Norman, R., Brazier, J., Cronin, P., King, M. T., Ratcliffe, J., & Street, D. (2014). An Australian discrete choice experiment to value EQ-5D health states. Health Economics, 23(6), 729-742.
14. Brazier, J., Czoski-Murray, C., Roberts, J., Brown, M., Symonds, T., & Kelleher, C. (2008). Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. Medical Decision Making, 28(1), 113-126.
15. Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 Health Survey. Journal of Clinical Epidemiology, 51(11), 1115-1128.
16. Young, T. A., Yang, Y., Brazier, J. E., & Tsuchiya, A. (2011). The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. Medical Decision Making, 31(1), 195-210.
17. Young, T. A., Yang, Y., Brazier, J. E., Tsuchiya, A., & Coyne, K. (2009). The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. Quality of Life Research, 18(2), 253-265.
18. Norman, R., Viney, R., Aaronson, N. K., Brazier, J. E., Cella, D. F., Costa, D. S. J., Fayers, P. M., Kemmler, G., Pickard, A. S., Rowen, D., Street, D., Velikova, G., Young, T. A., & King, M. T. (2015). Using a discrete choice experiment to value the QLU-C10D: Feasibility and sensitivity to presentation format. Quality of Life Research, Published online 5 Sept 2015 (DOI 10.1007/s11136-015-1115-3).
19. Pickard, A. S., Shaw, J. W., Lin, H. W., Trask, P. C., Aaronson, N., Lee, T. A., & Cella, D. (2009). A patient-based utility measure of health for clinical trials of cancer therapy based on the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire. Value in Health, 12(6), 977-988.
20. Bjordal, K., de Graef, A., Fayers, P. M., Hammerlid, E., van Pottelsberghe, C., Curran, D., Ahlner-Elmqvist, M., Maher, E. J., Meyza, J. W., Bredart, A., Soderholm, A. L., Arraras, J. J., Feine, J. S., Abendstein, I. H., Morton, R. P., Pignon, T., Huguenin, P., Bottomly, A., Kaasa, A., & Group, o. b. o. t. E. Q. o. L. (2000). A 12 country field study of the EORTC QLQ-C30 (version 3.0) and the head and neck cancer specific module (EORTC QLQ-H&N35) in head and neck patients. European Journal of Cancer, 36, 1796-1807.

21. Costa, D. S. J., Aaronson, N. K., Fayers, P. M., Grimison, P., Janda, M., Pallant, J. F., Rowen, D. L., Velikova, G., Viney, R., Young, T. A., & King, M. T. (2014). Deriving a preference-based utility measure for cancer patients from the EORTC QLQ-C30: a confirmatory versus exploratory approach. Patient Related Outcome Measures, 5, 119-129.
22. Costa, D. S. J., Aaronson, N. K., Fayers, P. M., Pallant, J. F., Velikova, G., & King, M. T. (2015). Testing the measurement invariance of the EORTC QLQ-C30 across primary cancer sites using multi-group confirmatory factor analysis. Quality of Life Research, 24(1), 125-133.
23. Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). British Journal of Clinical Psychology, 46, 1-18.
24. Revicki, D., Hays, R. D., Cella, D., Sloan, J., Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. Journal of Clinical Epidemiology, 61(2), 102-109.
25. Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. Medical Care, 27(3 Suppl), S178-S189.
26. Costa, D. S. J., King, M. T., Aaronson, N. K., Brazier, J. E., Cella, D., Grimison, P., Janda, M., McTaggart-Cowan, H., Pickard, A. S., Rowen, D., & Di Sipio, T. (2014). Do cancer patients distinguish between the importance and severity of their quality of life concerns? Quality of Life Research, 23(1 Supplement), 122.
27. Arraras, J. I., Suárez, J., Arias de la Vega, F., Vera, R., Asín, G., Arrazubi, V., Rico, M., Teijeira, L., & Azparren, J. (2011). The EORTC Quality of Life questionnaire for patients with colorectal cancer: EORTC QLQ-CR29 validation study for Spanish patients. Clinical and Translational Oncology, 13(1), 50-56.
28. Arraras, J. I., Tejedor, M., Illaramendi, J. J., Vera, R., Pruja, E., Marcos, M., Arias, F., & Valerdi, J. J. (2001). El cuestionario de calidad de vida para cáncer de mama de la EORTC, QLQ-BR23: Estudio psicométrico con una muestra española. Psicologia Conductual, 9, 81-97.
29. Arraras Urdaniz, J. I., Villafranca Iturre, E., Arias de la Vega, F., Domínguez Domínguez, M. A., Lainez Milagro, N., Manterola Burgaleta, A., Martínez Lopez, E., Romero Rojano, P., & Martínez Aguillo, M. (2008). The EORTC Quality of Life Questionnaire QLQ-C30 (version 3.0). Validation study for Spanish prostate cancer patients. Archivos Españoles de Urología, 61(8), 949-954.
30. Cengiz, M., Ozyar, E., Esassolak, M., Altun, M., Akmansu, M., Sen, M., Uzel, O., Yavuz, A., Dalmaz, G., Uzal, C., Hiçsönmez, A., Sarihan, S., Kaplan, B., Atasoy, B. M., Ulutin, C., Abacioğlu, U., Demiral, A. N., & Hayran, M. (2005). Assessment of quality of life of nasopharyngeal carcinoma patients with EORTC QLQ-C30 and H&N-35 modules. International Journal of Radiation Oncology Biology Physics, 63(5), 1347-1353.
31. Chie, W., Chang, K., Huang, C., & Kuo, W. (2003). Quality of life of breast cancer patients in Taiwan: validation of the Taiwan Chinese version of the EORTC QLQ-C30 and EORTC QLQ-BR23. Psycho-Oncology, 12, 729-735.

32. Clarke, S. J., Yip, S., Brown, C., van Hazel, G. A., Ransom, D. T., Goldstein, D., Jeffrey, G. M., Tebbutt, N. C., Buck, M., Lowenthal, R. M., Boland, A., Gebski, V., Zalcberg, J., & Simes, R. J. (2011). Single-agent irinotecan or FOLFIRI as second-line chemotherapy for advanced colorectal cancer; results of a randomised phase II study (DaVINCI) and meta-analysis. European Journal of Cancer, 47(12), 1826-1836.
33. Kaasa, S., Brenne, E., Lund, J. A., Fayers, P., Falkmer, U., Holmberg, M., Lagerlund, M., & Bruland, O. (2006). Prospective randomised multicenter trial on single fraction radiotherapy (8 Gy x1) versus multiple fractions (3 Gy x10) in the treatment of painful bone metastases. Radiotherapy and Oncology, 29, 278-284.
34. Mystakidou, K., Tsilika, E., Parpa, E., Kalaidopoulou, O., Smyrniotis, V., & Vlahos, L. (2001). The EORTC core quality of life questionnaire (QLQ-C30, version 3.0) in terminally ill cancer patients under palliative care: Validity and reliability in a Hellenic sample. Internation Journal of Cancer, 94(1), 135-139.
35. Richardson, P. G., Barlogie, B., Berenson, J., Singhal, S., Jagannath, S., Irwin, D., Rajkumar, S. V., Srkalovic, G., Alsina, M., Alexanian, R., Siegel, D., Orlowski, R. Z., Kuter, D., Limentani, S. A., Lee, S., Hideshima, T., Esseltine, D., Kauffman, M., Adams, J., Schenkein, D. P., & Anderson, K. C. (2003). A phase 2 study of Bortezomib in relapsed, refractory myeloma. New England Journal of Medicine, 348(2609-2617).
36. Singer, S., Wollbrück, D., Wulke, C., Dietz, A., Klemm, E., Oeken, J., Meister, E. F., Gudziol, H., Bindewald, J., & Schwarz, R. (2009). Validation of the EORTC QLQ-C30 and EORTC QLQ-H&N35 in patients with laryngeal cancer after surgery. Head & Neck, 31(1), 64-76.
37. Smith, A. B., King, M. T., Butow, P., Luckett, T., Grimison, P., Toner, G. C., Stockler, M., Hovey, E., Stubbs, J., Hruby, G., Gurney, H., Turner, S., Alam, M., Cox, K., & Olver, I. (2013). The prevalence and correlates of supportive care needs in testicular cancer survivors: A cross-sectional study. Psycho-Oncology, 22(11), 2557-2264.
38. Tebbutt, N. C., Cummins, M. M., Sourjina, T., Strickland, A. H., Van Hazel, G., Ganju, V., Gibbs, D., Stockler, M., Gebski, V., Zalcberg, J., & on behalf of the Australasian Gastro-Intestinal Trials Group. (2010). Randomised, non-comparative phase II study of weekly docetaxel with cisplatin and 5-fluorouracil or with capecitabine in oesophagogastric cancer: the AGITG ATTAX trial. British Journal of Cancer, 102, 475–481.
39. Tebbutt, N. C., Parry, M. M., Zannino, D., Strickland, A. H., Van Hazel, G. A., Pavlakis, N., Ganju, V., Mellor, D., Dobrovic, A., Gebski, V. J., & on behalf of the Australasian Gastro-Intestinal Trials Group (AGITG). (2013). Docetaxel plus cetuximab as second-line treatment for docetaxel-refractory oesophagogastric cancer: the AGITG ATTAX2 trial. British Journal of Cancer, 108(4), 771-774.
40. Teckle, P., Peacock, S., McTaggart-Cowan, H., van der Hoek, K., Chia, S., Melosky, B., & Gelmon, K. (2011). The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities Health and Quality of Life Outcomes, 9, 106.

41. Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J., & Selby, P. J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. Journal of Clinical Oncology, 22(4), 714-724.
42. Velikova, G., Sheppard, S., Campbell, L., Smith, A., Awad, N., Selby, P., & Brown, J. (2008). Randomized trial of quality-of-life measurement in oncology practice: Do oncologists need to know? , ASCO Annual Meeting (Vol. 26, pp. 9586). Chicago: Journal of Clinical Oncology.
43. Whistance, R. N., Conroy, T., Chie, W., Costantini, A., Sezer, O., Koller, M., Johnson, C. D., Pilkington, S. A., Arraras, J., Ben-Josef, E., Pullyblank, A. M., Fayers, P., Blazeby, J. M., & on behalf of the European Organisation for the Research and Treatment of Cancer (EORTC) Quality of Life Group. (2009). Clinical and psychometric validation of the EORTC QLQ-CR29 questionnaire module to assess health-related quality of life in patients with colorectal cancer. European Journal of Cancer, 45, 3017-3026.
44. Klepstad, P., Dale, O., Kaasa, S., Zahlsen, K., Aamo, T., Fayers, P., & Borchgrevink, P. C. (2003). Influences on serum concentrations of morphine, M6G and M3G during routine clinical drug monitoring: A prospective survey in 300 adult cancer patients. Acta Anaesthesiologica Scandinavica, 47, 725-731.
45. Sintonen, H. (2001). The 15D instrument of health-related quality of life: properties and applications. Annals of Medicine, 33, 328-336.
46. Rowen, D., Brazier, J. E., Young, T., Gaugris, S., Craig, B. M., King, M. T., & Velikova, G. (2011). Deriving a preference-based measure for cancer using the EORTC-QLQC30. Value in Health, 14(5), 721-731.
47. Costa, D. S., Aaronson, N. K., Fayers, P. M., Grimison, P., Janda, M., Pallant, J. F., Rowen, D. L., Velikova, G., Viney, R., Young, T. A., & King, M. T. (In press (accepted 1st Aug 2014)). Deriving a preference-based utility measure for cancer patients from the EORTC QLQ-C30: a confirmatory versus exploratory approach. Patient Related Outcome Measures.
48. Costa, D. S. J., King, M. T., Aaronson, N. K., Brazier, J. E., Cella, D., Grimison, P., Janda, M., McTaggart-Cowan, H., Pickard, A. S., Rowen, D., & Di Sipio, T. (2014). Do cancer patients distinguish between the importance and severity of their quality of life concerns?, Annual Conference for the International Society for Quality of Life Research (Vol. 23, pp. 122). Berlin, Germany: Quality of Life Research.
49. McTaggart-Cowan, H., Teckle, P., & Peacock, S. (2013). Mapping utilities from cancer-specific health-related quality of life instruments: A review of the literature. Expert Review of Pharmacoeconomics & Outcomes Research, 13(6), 753-765.
50. Fayers, P. M., & Hays, R. D. (2014). Should linking replace regression when mapping from profile-based measures to preference-based measures? Value in Health, 17, 261-265.
51. Brazier, J. E., Yang, Y., Tsuchiya, A., & Rowen, D. L. (2010). A review of studies mapping (or cross walking) non preference based measures of health to generic preference-based measures. The European Journal of Health Economics, 11, 215-225.
52. Scott, N. W., Fayers, P., Aaronson, N., Bottomley, A., de Graeff, A., Groenvold, M., Gundy, C., Koller, M., Peterson, M. A., Sprangers, M. A. G., &

Group, o. b. o. t. E. Q. o. L. (2008). EORTC QLQ-C30 Reference Values. In E. Q. o. L. Department (Ed.). Brussels: European Organisation for Research and Treatment of Cancer (EORTC).
53. Jim, H. S. L., Phillips, K. M., Chait, S., Faul, L. A., Popa, M. A., Lee, Y. H., Hussin, M. G., Jacobsen, P. B., & Small, B. J. (2012). Meta-analysis of cognitive functioning in breast cancer survivors previously treated with standard-dose chemotherapy. Journal of Clinical Oncology, 30(29), 3578-3587.
54. Park, J. H., & Bae, S. H. (2012). A meta-analysis of chemotherapy related cognitive impairment in patients with breast cancer. Journal of Korean Academy of Nursing, 42(5), 644-658.

Table 1 Characteristics of the 18 datasets and the number of observations each contributed to various psychometric analyses.

PI of study [Reference number]	Study design [1]	Primary cancer site and stage	Country	Treatment[2]				Obs/ patient[3]	n[4] CFA/ Rasch	n[4] Respons-iveness
				Chemo†	Radio†	Surgery†	Other†			
Abacioglu [31]	Obs	mixed sites and stages	Turkey	12	88	0	0	2+	88	103-104
Arraras 1 [30]	Val	prostate, early stage	Spain	0	117	0	0	1+	117	96-97
Arraras 2 [29]	Val	breast, early stage	Spain	33	68	93	53	1+	94	-
Arraras 3 [28]	Val	colorectal, late stage	Spain	44	0	0	0	1+	44	26-28
Blazeby [44]	Val	colorectal, mixed stages	UK, US, Taiwan	189	61	149	29	2+	277	-
Brenne [34]	RCT	mixed sites, late stage	Norway, Sweden	0	349	0	0	2+	349	134-136
Chie 1 [32]	Obs	breast, early stage	Taiwan	33	7	0	55	1	84	-
Clarke [33]	RCT	colorectal, metastatic	Australia	74	0	0	0	2+	74	62-68
Klepstad [45]	Obs	mixed sites, late stage	Norway	0	0	0	183	1	183	-
Millennium Pharmaceuticals [36]	Obs	multiple myeloma, late stage	UK	187	0	0	0		187	122-127
Mystakidou [35]		mixed sites, in palliative care	Greece	-	-	-	-	2+	-	60-62
Olver [38]	Obs	testicular, mixed stages	Australia	149	67	242	5	1	270	-
Peacock [41]	Obs	mixed sites and stages	Canada	266	40	3	21	1	301	-
Schwarz 3 [37]	Obs	head & neck, early stage	Germany	0	0	117	0	1	117	-
Tebbutt 1 [39]	RCT	oesophago-gastric, advanced	Australia	67	0	0	0	2+	67	91-94
Tebbutt 2 [40]	RCT	oesophago-gastric, advanced	Australia	35	0	0	0	2+	35	*
Velikova 1 [43]	RCT	mixed sites and stages	UK	138	0	0	0	2+	138	33
Velikova 2 [42]	RCT	mixed sites and stages	UK	191	0	0	0	2+	191	*
Subtotal									**2616**	**727-749**

1. Observational clinical study (Obs); randomised controlled trial (RCT); EORTC QOL module validation studies (Val).
2. Treatment categories are not mutually exclusive
3. HRQOL assessments per patient categorised as: 1; 1 or more (1+); 2 or more (2+)
4. Number of patients (n) each dataset contributed for CFA/Rasch and responsiveness analyses.

* The two Tebbutt data sets were pooled for responsiveness analyses, as were the two Velikova data sets.

Table 2 Summary of results for the nine criteria by which items were assessed for suitability to represent their respective dimension within the health classification system of the multi-attribute utility instrument. Items in italics were those selected for the HSCS.

Criteria[a]			1	2		3	3	4	6	7	8	9		
QLQ-C30 domain	Item number and topic	loading	Fit	DT	Loc-ation	Lowest threshold[b]	Highest threshold[b]	DIF	Floor Ceiling	Cohen's d	Tally[c] (top)	Importance rating group[d]	Most Import-ant[f]	Top 3
Physical functioning	1. Trouble doing strenuous activities	0.8	-1.1		-1.82	-1.85	1.63			0.58	4 (2)	Physical functioning	34	68
	2. Trouble taking a long walk	0.85	-1.35		-1.63	-1.81	1.73			0.75	5 (2)		19	63
	3. Trouble taking a Short walk	0.85	-2.02		0.96	-1.60	1.94			0.68	6 (1)		4	25
	4. Need to stay in bed/chair	0.83	1.86		-0.03	-1.82	1.64			0.80	6 (2)		10	36
	5. Need help eating, dressing, washing, using toilet	0.64	0.12	X	2.52	-1.09	1.09		Floor	0.49	4 (2)		3	3
Emotional functioning	21. Tense	0.86	0.49		-0.22	-2.65	1.96			0.23	6 (2)	Emotional functioning	6	51
	22. Worry[e]	0.85	-1.09		-0.6	-2.48	2.42			0.23	2 (0)		41	65
	23. Irritable	0.78	0.45		0.41	-2.67	2.72			0.16	6 (3)		13	47
	24. Depressed[e]	0.84	0.83		0.41	-2.49	2.54			0.26	5 (2)		11	49
Role Functioning	*6. Limited in doing work / other daily activities*	0.93	0.48		0.08	-2.21	2.13	site		0.78	6 (3)	Other functioning	36	72
	7. Limited in doing hobbies / other leisure activities	0.91	0.36		-0.08	-2.04	2.07	site		0.75	6 (4)		13	42
Social Functioning	*26. Physical condition/medical treatment interfered with family life*	0.82	0.91		0.52	-2.76	2.43			0.53	4 (3)	Other functioning	13	39
	27. Physical condition/medical treatment interfered with social activities	0.89	-0.18		-0.52	-2.00	2.00			0.62	5 (3)		4	27
Cognitive Functioning	20. Difficulty concentrating	0.87	0.06		-0.04	-0.59	1.10			0.76	5 (2)	Other functioning	4	21
	25. Difficulty remembering things	0.69	0.79		0.04	-1.96	1.46			0.39	5 (3)		7	34
Fatigue	10. Need to rest	0.89	-0.56		-0.31	-3.09	2.63			0.69	7 (2)	Symptoms	12	67
	12. Felt weak	0.87	0.12		0.47	-2.69	2.11			0.61	7 (2)		15	51
	18. Tired	0.83	-0.01		-0.16	-3.34	3.08			0.58	7 (3)		27	73
Pain	*9. Had pain*	0.86	0.39		-0.41	-4.46	3.97			0.66	3 (1)	Symptoms	9	18
	19. Pain interfered with daily activities	0.94	-0.42		0.41	-2.88	1.91			0.62	4 (3)		8	16

Nausea/ vomiting	*14. Felt nauseated*	0.88	-0.64		-1.68	-2.73	2.63			0.55	6 (4)	Symptoms	5	22
	15. Vomited	0.82	0.05		1.68	-3.37	5.65		Floor	0.48	4 (2)	Symptoms	2	3
Single-item symptoms	8. Short of breath (Dyspnoea)	-	-	-	-			-		0.53	5 (1)	Other difficulties	7	35
	11. Trouble sleeping	-	-	-	-			-		0.23	3 (1)		24	56
	13. Lacked appetite	-	-	-	-			-		0.63	9 (3)		11	28
	16. Constipated	-	-	-	-			-		0.51	7 (1)		13	39
	17. Diarrhoea	-	-	-	-			-		0.24	7 (3)		7	17

a. Criteria: 1) fit of items to the Rasch model; 2) disordered response thresholds (DT); 3) spread of item thresholds across the latent variable; 4) differential item function (DIF); 5) Local dependence; 6) floor and ceiling effects; 7) sensitivity to differences between early and late stage cancer; 8) responsiveness to change due to treatment: 9) Patient opinion about relative importance of items within domains. Criterion 5 is not shown in this table as none of the items showed local dependence within their respective domains.

b. Criterion 3 – Spread of item thresholds across the latent variable: These thresholds are for the model with Item 5 rescored as described in text.

c. Criterion 8 - Responsiveness: Tally = number of effect size estimates of at least 0.20; Top = number of times the item was the most responsive.

d. Criterion 9 - Patient Importance rating groups differed from QLQ-C30 domains for those dimensions that contains two or less items; these were collated into groups of 5-6 items to allow patients to identify the Top 3 issues that most affected their quality of life. The frequency with which each item was nominated as the Most Important and in the Top 3 is tabulated.

e. While the Worry item was initially chosen, following our methodological study to determine a valid and feasible valuation method to generate a preference-based utility scoring algorithm, the final version of the QLU-C10D reverted to the Depression item, for reasons described in the companion paper by Norman et al.

f. In principle the “Most important” column should sum to the sample size, and “Top 3” should sum to triple the sample size. Because some participants nominated more than the requested number of items, and some fewer, the values in this table deviate from what was expected.

Table 3 The dimensions and levels of the QLU-C10D health state classification system derived from the EORTC core quality of life questionnaire, QLQ-C30.

Dimension	QLQ-C30 Item (s)	Level 1	Level 2	Level 3	Level 4
Physical functioning	2, 3	I have no trouble taking a long walk	I have at least a little trouble taking a long walk but no trouble taking a short walk outside the house	I have a little trouble taking a short walk outside the house	I have quite a bit or very much trouble taking a short walk outside the house
Role functioning	6	I am not limited in pursuing my work or other daily activities	I am limited a little in pursuing my work or other daily activities	I am limited quite a bit in pursuing my work or other daily activities	I am limited very much in pursuing my work or other daily activities
Social functioning	26, 27	My physical condition or medical treatment does not interfere with my social or family life	My physical condition or medical treatment interferes with my social or family life a little	My physical condition or medical treatment interferes with my social or family life quite a bit	My physical condition or medical treatment interferes with my social or family life very much
Emotional functioning	24	I do not feel depressed	I feel a little depressed	I feel quite a bit depressed	I feel very much depressed
Pain	9	I do not have pain	I have a little pain	I have pain quite a bit	I have pain very much
Fatigue	18	I do not feel tired	I feel a little tired	I feel tired quite a bit	I feel tired very much
Sleep	11	I do not have trouble sleeping	I have a little trouble sleeping	I have quite a bit of trouble sleeping	I have very much trouble sleeping
Appetite	13	I do not lack appetite	I lack appetite a little	I lack appetite quite a bit	I lack appetite very much
Nausea	14	I do not feel nauseated	I feel a little nauseated	I feel nauseated quite a bit	I feel nauseated very much
Bowel problems	16, 17	I do not have constipation or diarrhoea	I have a little constipation or diarrhoea	I have constipation or diarrhoea quite a bit	I have constipation or diarrhoea very much

Online Appendix - Figures

Figure A. Item map for Physical Functioning (with Item 5 rescored), items 1-5. Item maps show the distributions of Rasch-derived latent variable scores (location parameters) for participants and items, illustrating to what extent the levels of the trait captured by each item matches the level of the trait for each participant.

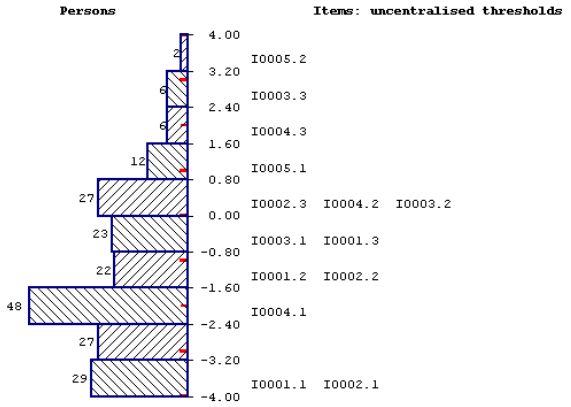


Figure B. Item map for Role Functioning, items 6-7

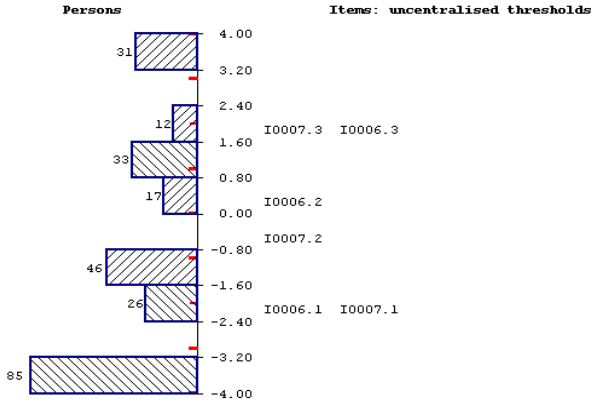


Figure C. Item map for Emotional Functioning, items 21-24

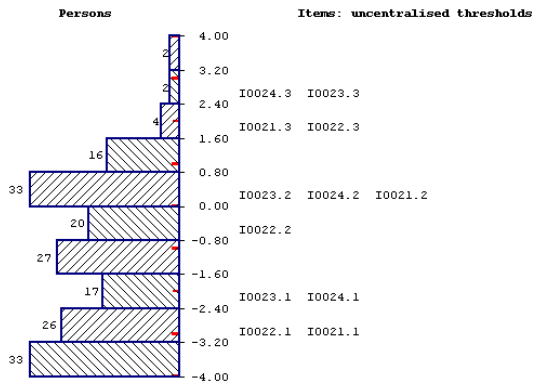


Figure D. Item map for Social Functioning, items 26, 27

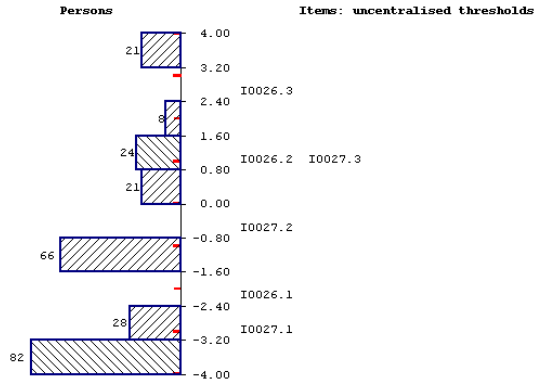


Figure E. Item map for Cognitive Functioning, items 20, 25

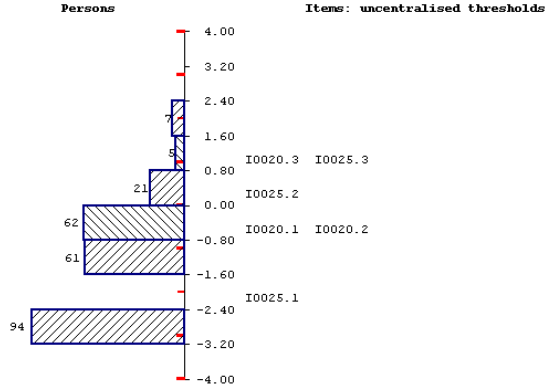


Figure F. Item map for Pain, items 9, 19

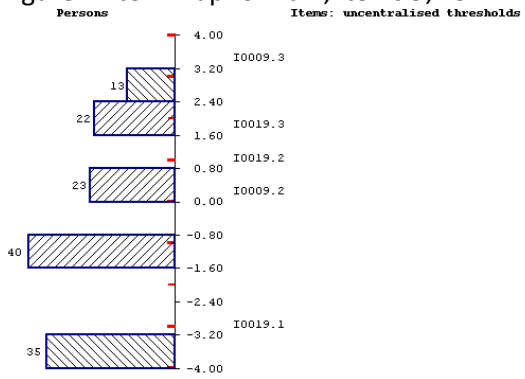


Figure G. Item map for Fatigue, items 10, 12, 18

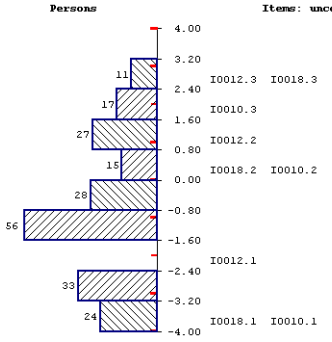


Figure H. Item map for Nausea and Vomiting, items 14, 15

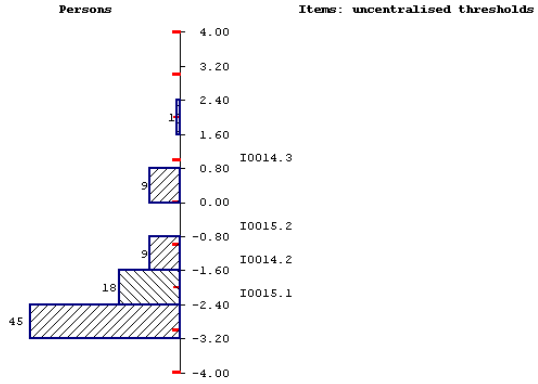
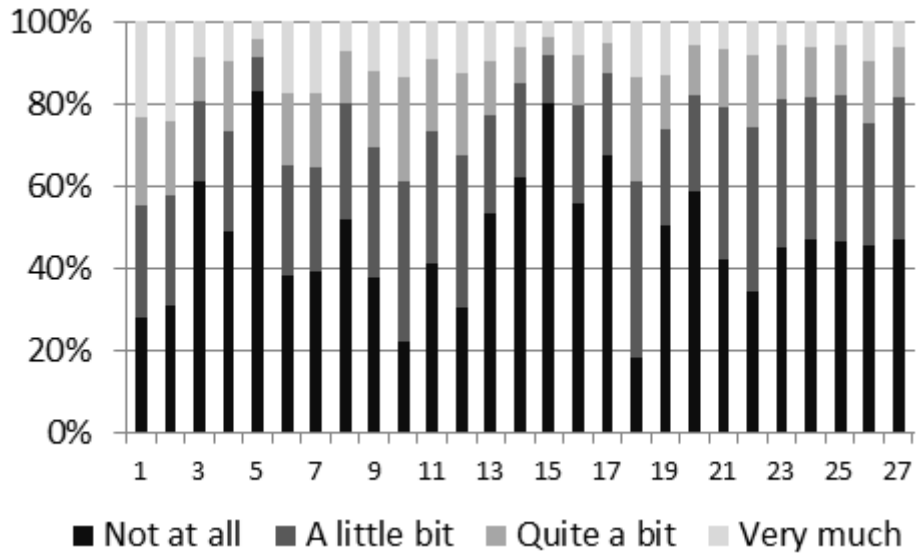


Figure I. Response frequencies for Items 1-27 items of the QLQ-C30 used to assess Criterion 6 (floor and ceiling effects)



Online Appendix – Tables A-D

Table A. Frequency of observations by primary cancer site and stage for the pooled dataset used to assess Criteria 1-7 (n=2616 patients)

	Loco-regional		Recurrent/metastatic	
	Frequency	**Percent**	**Frequency**	**Percent**
Breast	287	27.7	250	15.8
Colorectal	240	23.1	262	16.6
Genito-urinary	9	0.9	47	3.0
Gynaecological	33	3.2	95	6.0
Head and neck	119	11.5	2	0.1
Leukaemia	0	0.0	17	1.1
Hepato-biliary (liver/bile/pancreas)	0	0.0	4	0.3
Lung	40	3.9	158	10.0
Melanoma	1	0.1	1	0.1
Myeloma	0	0.0	191	12.1
Gastro-oesophageal (oesophagus/stomach)	3	0.3	121	7.7
Prostate	119	11.5	177	11.2
Sarcoma	7	0.7	33	2.1
Testicular	176	17.0	106	6.7
Other	3	0.3	115	7.3
Total	**1037**		**1579**	

Table B. Frequency of observations for each treatment for the pooled dataset used to assess Criteria 1-7 (n=2616 patients)

	Frequency	**Percent**
No treatment	42	1.6
Chemotherapy only	1083	41.4
Radiotherapy only	569	21.8
Hormone therapy only	59	2.3
Surgery only	237	9.1
Analgesics only	183	7
Chemo + radiotherapy	65	2.5
Hormone + radiotherapy	2	0.1
Surgery + radiotherapy	77	2.9
Surgery + chemotherapy	195	7.5
Surgery + chemo + radiotherapy	42	1.6
Surgery + hormone therapy	8	0.3
Chemo + hormone therapy	8	0.3
Surgery + chemo + hormone therapy	5	0.2
Surgery + hormone + radiotherapy	24	0.9
Chemo + hormone + radiotherapy	1	0
Surgery + chemo + hormone + radiotherapy	16	0.6
Total	**2616**	

Table C. Criterion 8: Responsiveness to change (effect size[1]), by item and QLQ-C30 domain scale, for each study

Domain/Item		Tally[2] (Top[3])	Abac-ioglu [39]	Arra-ras1 [30]	Arra-ras3[4] [28]	Tebbutt[5] [39, 40]	Brenne [34]	Clarke [33]	Millenium [36]	Mysta-kidou [35]	Velikova[5] [42,43]
Physical Functioning			*0.19*	*0.19*	*0.55*	*0.73*	*0.96*	*0.32*		*-1.11*	
1	strenuous	4 (2)			0.29	0.42			**0.20**	**-1.25**	
2	long walk	5 (2)			**0.53**	0.59		0.32		-0.64	**0.25**
3	short walk	6 (1)	**0.25**	0.27	-	0.67	0.90	0.34		-0.61	
4	bed/chair	6 (2)		**0.45**	-0.28	0.44	0.61	**0.56**		-1.09	
5	help	4 (2)	0.20		-	**1.51**	**2.78**			-0.41	
Role Functioning			*0.08*	*0.26*	*0.48*	*0.52*	*0.27*	*0.25*		*-1.25*	
6	limited work	6 (3)		**0.29**	0.30	0.46	0.24	**0.25**		**-1.08**	
7	limited hobbies	6 (4)			**0.50**	**0.54**	**0.29**	0.22		-0.95	**-0.23**
Emotional Functioning			*-0.11*	*0.27*	*0.37*	*0.14*	*1.71*	*0.00*	*-0.14*	*-0.92*	*-0.25*
21	tense	6 (2)		0.24	**0.44**	0.22	1.42			-0.28	**-0.27**
22	worry	2 (0)					1.19			-0.69	
23	irritable	6 (3)		**0.37**	0.31	**0.25**	**1.83**			-0.60	-0.26
24	depressed	5 (2)	**-0.21**	0.23	0.26		1.35			**-0.94**	
Social Functioning			*-0.08*	*0.45*	*0.46*	*0.50*	*1.05*	*0.03*	*-0.01*	*-0.46*	*-0.22*
26	family	4 (3)		**0.61**		0.39	**1.05**				**-0.39**
27	social	5 (3)		0.27	**0.70**	**0.48**	0.74			**-0.49**	
Cognitive Functioning			*-0.03*	*0.51*	*0.42*	*0.12*	*1.80*	*0.37*	*-0.03*	*-0.42*	*-0.16*
20	concentrate	5 (2)		0.22	0.34		1.49	**0.41**		**-0.41**	
25	memory	5 (3)		**0.59**	**0.51**		**1.64**	0.30		-0.31	
Fatigue			*0.10*	*0.65*	*0.58*	*0.36*	*1.14*	*0.38*	*0.12*	*-1.48*	*-0.22*
10	rest	7 (2)		0.49	**0.86**	0.29	0.50	**0.42**		-1.20	-0.24
12	weak	7 (2)	**0.21**	0.64	0.34	0.32	**0.78**	0.34		-1.81	
18	tired	7 (3)		**1.03**	0.62	0.33	0.68	0.37		**-1.95**	**-0.25**
Pain			*0.06*	*0.43*	*0.49*	*0.02*	*0.26*	*0.07*	*0.03*	*-2.82*	*0.06*
9	pain	3 (1)		0.29	0.22					**-2.60**	
19	pain interfere	4 (3)		**0.50**	**0.81**		**0.28**			-1.93	
Nausea/vomiting			*0.14*	*0.12*	*1.74*	*0.08*	*2.16*	*0.80*	*0.28*	*-0.62*	*-0.27*
14	nausea	6 (4)			**0.73**		1.53	0.71	**0.36**	**-0.72**	**-0.29**
15	vomiting	4 (2)			-		**2.60**	**0.76**		-0.37	-0.20
Other symptoms											
8	dyspnoea	5 (1)		0.25		**0.33**	1.31	0.17		-0.47	
11	sleep	3 (1)			0.51		1.23			**-1.26**	
13	appetite	9 (3)	**0.29**	0.31	0.66	0.20	1.10	0.40	**0.25**	-0.73	**-0.27**
16	constipation	7 (1)		0.70	**0.91**	-0.24	0.97	0.59		-0.77	-0.26
17	diarrhoea	7 (3)	0.23	**1.10**	0.31	0.24	**2.76**	**0.69**		-0.29	

1. Responsiveness characterised as Effect Size = on-treatment minus baseline divided by baseline standard deviation. Scale direction standardised so that +'ve change always indicates deterioration whether symptom or functioning. Calculated for items 1-27 of the QLQ-C30, by study. Effect sizes for domain scales shown in italics. Effect sizes for items not shown if less than 0.20. Largest effect size per domain bolded; if next biggest is within 0.1 or is ≥ 1.0, it is underlined.
2. Tally = number of effect size estimates of at least 0.20.
3. Top = number of times the item was the most responsive
4. For Arraras 3, items 3, 5 and 15 did not produce a responsiveness index because there was no variance in the baseline measures – specifically, all patients responded "Not at all".
5. The two Tebbutt data sets were pooled, as were the two Velikova data sets.

Table D. Patient illness and treatment characteristics of patients in the two parts of the patient opinion study: survey development (qualitative phase, n=9), quantitative survey (n=79, used to assess Criterion 9).

		Qualitative (n=9)	Quantitative (n=79)
Age	Mean (SD)	56.7 (14.5)	60.0 (13.1)
Sex	Male	7	29
	Female	2	50
Primary cancer site	Colorectal	1	25
	Breast		17
	Ovarian		7
	Lung	1	3
	Oesophageal/Stomach	3	2
	Pancreatic	2	8
	Prostate	1	2
	Testicular	1	5
	Other		13
Stage	Localised	3	22
	Metastasised	4	39
	Not recorded	2	2
	Unknown		16
Treatment	Surgery	4	56
	Chemotherapy	8	76
	Radiotherapy	3	29
	Hormone therapy	2	17
	Other	2	
Time since diagnosis	< 6 months	6	51
	6-12 months	1	26
	12-24 months		8
	2-5 years	0	13
	5-10 years	2	17
	> 10 years		10
	Not recorded		5
Current treatment	Chemotherapy only	4	70
	Chemotherapy and radiotherapy	2	
	Hormone therapy only	1	1
	Chemotherapy and hormone		1
	None	1	
	Not recorded	1	7