Patient Related Outcome Measures

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Deriving a preference-based utility measure for cancer patients from the European Organisation for the Research and Treatment of Cancer's Quality of Life Questionnaire C30: a confirmatory versus exploratory approach

This article was published in the following Dove Press journal:
Patient Related Outcome Measures
6 November 2014
Number of times this article has been viewed

Daniel SJ Costa[1]
Neil K Aaronson[2]
Peter M Fayers[3,4]
Peter S Grimison[5,6]
Monika Janda[7]
Julie F Pallant[8]
Donna Rowen[9]
Galina Velikova[10]
Rosalie Viney[11]
Tracey A Young[9]
Madeleine T King[1]

On behalf of the MAUCa Consortium

[1]Psycho-oncology Co-operative Research Group, University of Sydney, Sydney, NSW, Australia; [2]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, the Netherlands; [3]Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK; [4]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway; [5]Chris O'Brien Lifehouse, [6]Sydney Medical School, University of Sydney, Sydney, NSW, [7]School of Public Health, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, [8]Rural Health Academic Centre, University of Melbourne, Shepparton, VIC, Australia; [9]School of Health and Related Research, University of Sheffield, Sheffield; [10]University of Leeds, St James's Institute of Oncology, Leeds, UK; [11]Centre for Health Economics Research and Evaluation, University of Technology, Sydney, NSW, Australia

Correspondence: Daniel SJ Costa
Psycho-oncology Co-operative Research Group, Chris O'Brien Lifehouse (C39Z), University of Sydney, Sydney, NSW 2006, Australia
Tel +61 2 9351 6304
Fax +61 2 9036 5292
Email daniel.costa@sydney.edu.au

Background: Multi attribute utility instruments (MAUIs) are preference-based measures that comprise a health state classification system (HSCS) and a scoring algorithm that assigns a utility value to each health state in the HSCS. When developing a MAUI from a health-related quality of life (HRQOL) questionnaire, first a HSCS must be derived. This typically involves selecting a subset of domains and items because HRQOL questionnaires typically have too many items to be amendable to the valuation task required to develop the scoring algorithm for a MAUI. Currently, exploratory factor analysis (EFA) followed by Rasch analysis is recommended for deriving a MAUI from a HRQOL measure.

Aim: To determine whether confirmatory factor analysis (CFA) is more appropriate and efficient than EFA to derive a HSCS from the European Organisation for the Research and Treatment of Cancer's core HRQOL questionnaire, Quality of Life Questionnaire (QLQ-C30), given its well-established domain structure.

Methods: QLQ-C30 (Version 3) data were collected from 356 patients receiving palliative radiotherapy for recurrent/metastatic cancer (various primary sites). The dimensional structure of the QLQ-C30 was tested with EFA and CFA, the latter informed by the established QLQ-C30 structure and views of both patients and clinicians on which are the most relevant items. Dimensions determined by EFA or CFA were then subjected to Rasch analysis.

Results: CFA results generally supported the proposed QLQ-C30 structure (comparative fit index =0.99, Tucker–Lewis index =0.99, root mean square error of approximation =0.04). EFA revealed fewer factors and some items cross-loaded on multiple factors. Further assessment of dimensionality with Rasch analysis allowed better alignment of the EFA dimensions with those detected by CFA.

Conclusion: CFA was more appropriate and efficient than EFA in producing clinically interpretable results for the HSCS for a proposed new cancer-specific MAUI. Our findings suggest that CFA should be recommended generally when deriving a preference-based measure from a HRQOL measure that has an established domain structure.

Keywords: multi attribute utility instrument, health state classification system, confirmatory factor analysis, exploratory factor analysis, European Organisation for the Research and Treatment of Cancer QLQ-C30

Introduction

Multi attribute utility instruments (MAUIs) are preference-based quality of life measures that can be used in cost–utility analysis.[1] MAUIs have two components. The first

is a "health state classification system" (HSCS), comprising core domains of health-related quality of life (HRQOL), each comprising a number of levels (eg, poor, moderate, good). For example, the widely used MAUI, EQ-5D, has five dimensions, each with three levels.[2] These dimensions (or "attributes") and levels define the HSCS. Thus, the HSCS of the EQ-5D comprises 3^5=243 unique health states. The second component is a scoring algorithm, which assigns a utility value to each health state, based on the valuation elicited, using a preference-based assessment method, typically from a general population sample.

MAUIs have previously been derived from various HRQOL measures.[3–5] This typically involves two stages. The first stage involves selecting a subset of domains and items from the HRQOL measure to form a HSCS. This reduction stage is required because HRQOL measures typically include more items and domains than is manageable in the preference-based valuation exercise 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 European Organisation for Research and Treatment of Cancer's (EORTC) core Quality of Life Questionnaire (QLQ-C30)[6] is one of the most widely used cancer-specific HRQOL instruments, but is not a preference-based measure[7] and, therefore, cannot be used in cost–utility analysis. One solution is to "map" the QLQ-C30 to a preference-based measure.[8] A more theoretically rigorous approach is to develop a cancer-specific MAUI from the QLQ-C30, as has been done by Rowen et al.[9]

Rowen et al applied the methods of Young et al,[10] starting with exploratory factor analysis (EFA) to identify clusters of correlated items as a prerequisite to Rasch analysis to assess psychometric properties of items relevant to their performance in a MAUI.[5,10] Items that did not perform well on various psychometric criteria related to EFA and/or Rasch procedures were excluded, and then one or two items from each domain were retained as the basis for the HSCS for the MAUI. The main advantages of this method are that the resulting classification system represents the dimensionality of the measure using observed data. Further, this method can be used for any measures, regardless of whether it has an established dimensional structure. One crucial disadvantage is that EFA will produce only factors, as opposed to clinically coherent HRQOL dimensions.

When a HSCS is to be derived from a questionnaire with an established dimensional structure that is psychometrically robust and clinically sensible, arguably a confirmatory approach to the question of dimensionality is more appropriate than an exploratory approach. The QLQ-C30 is such an instrument. The confirmatory approach involves the positing of a specific dimensional structure (the conceptual model) that is tested with confirmatory factor analysis (CFA). This has three advantages over the exploratory approach. First, many of the arbitrary decisions involved in EFA (eg, method of extraction, method of rotation, number of factors to extract) are removed, replaced instead with more theoretically or clinically driven decisions, such as which items are hypothesized to load on which factors. Second, without a priori clinical guidance, any given solution may lack clinical cohesion. Third, the positing of a specific model allows clinical considerations – which we define here as the views of both patients and clinicians about issues relevant to HRQOL in cancer – to play a more structured a priori role than EFA can allow. Certain items may be included in or excluded from the model a priori, based on clinical or theoretical considerations, meaning that clinical considerations can be built in to the general method of item assessment, rather than acting as a post hoc, context-specific activity. Items deemed important in the trade-off between HRQOL and survival may thus be selected solely according to clinical considerations. For such items, clinical considerations would override statistical criteria, ensuring that the condition-specific preference-based measure contains symptoms of particular relevance to that condition. In cancer, these include fatigue, pain, and nausea.[11,12]

The aim of the current paper is to compare confirmatory with exploratory approaches in deriving a cancer-specific MAUI from the QLQ-C30, given its well-established domain structure. Note that the objective of the analyses reported in this paper was not to develop a specific HSCS, but rather to refine and make further recommendations on the appropriate methodology for defining the dimension structure for the MAUI, focusing on step 1 of the seven-step item selection procedure described by Young et al.[10]

Methods

Ethical approval for this study was granted by the University of Sydney Human Research Ethics Committee (Protocol Number 13207).

Quality of life instrument

The European Organisation for the Research and Treatment of Cancer QLQ-C30 (Version 3) is a multidimensional instrument containing 30 items assessing symptoms, functioning, and overall HRQOL (Table 1). Its validity and reliability are

Table 1 The 30 items of the Quality of Life Questionnaire C30 and the scales[a] to which they belong

Item	Item stem wording	Scale
1	Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	PF
2	Do you have any trouble taking a long walk?	PF
3	Do you have any trouble taking a short walk outside of the house?	PF
4	Do you need to stay in bed or a chair during the day?	PF
5	Do you need help with eating, dressing, washing yourself, or using the toilet?	PF
6	Were you limited in doing either your work or other daily activities?	RF
7	Were you limited in pursuing your hobbies or other leisure time activities?	RF
8	Were you short of breath?	Dyspnea (S)
9	Have you had pain?	Pain
10	Did you need to rest?	Fatigue
11	Have you had trouble sleeping?	Insomnia (S)
12	Have you felt weak?	Fatigue
13	Have you lacked appetite?	Appetite loss (S)
14	Have you felt nauseated?	Nausea/vomiting
15	Have you vomited?	Nausea/vomiting
16	Have you been constipated?	Constipation (S)
17	Have you had diarrhea?	Diarrhea (S)
18	Were you tired?	Fatigue
19	Did pain interfere with your daily activities?	Pain
20	Have you had difficulty in concentrating on things, like reading a newspaper or watching television?	CF
21	Did you feel tense?	EF
22	Did you worry?	EF
23	Did you feel irritable?	EF
24	Did you feel depressed?	EF
25	Have you had difficulty remembering things?	CF
26	Has your physical condition or medical treatment interfered with your family life?	SF
27	Has your physical condition or medical treatment interfered with your social activities?	SF
28	Has your physical condition or medical treatment caused you financial difficulties?	Financial difficulties (S)
29	How would you rate your overall health during the past week?	Global
30	How would you rate your overall quality of life during the past week?	Global

Notes: [a]There are five multi-item functioning scales (PF, RF, CF, EF, and SF); three multi-item symptom scales (fatigue, pain, and nausea/vomiting); a global health/quality of life scale; and six single-item scales (S).
Abbreviations: CF, cognitive functioning; EF, emotional functioning; PF, physical functioning; RF, role functioning; SF, social functioning.

well established.[6,13] Responses to items 1–28 are made on a four-point scale (1= "Not at all", 2= "A little", 3= "Quite a bit", 4= "Very much"), and responses to items 29 and 30 (global health and quality of life items) are made on a seven-point scale (1= "Very poor" and 7= "Excellent"). Items 6–30 have a recall period of the past week; no recall period is specified for items 1–5 (Physical Functioning). The 30 items form five functioning scales, three multi-item symptom scales, five single-item symptom scales (plus a financial difficulties item), and a global health status and HRQOL scale (Table 1).

Data set

A secondary analysis was conducted on data collected with the QLQ-C30 (Version 3) from a sample of 356 patients (53% Norwegian and 47% Swedish) with stage IV/recurrent/metastatic cancer from a variety of primary sites (36% prostate, 30% breast, 11% lung, and 23% other), all undergoing palliative radiotherapy in a randomized clinical trial comparing two fractionations.[14] The mean age was 66.77 years (standard deviation =10.60, range 31.59–90.32) and 43.8% were female. Analysis was conducted on the 316 of 356 patients who had complete QLQ-C30 data. These patients did not differ from those excluded on any of the key variables (assessed with chi-squared test for treatment arm [P=1.00], country [P=0.77], sex [P=0.06], and primary cancer site [P=0.72]).

Analysis

Exploratory versus confirmatory factor analysis

EFA is a statistical procedure in which variables are grouped into relatively independent subsets based on their intercorrelations, without any prior assumptions about the composition of these subsets. In contrast, CFA involves testing a prespecified arrangement of items into subsets, guided by a conceptual model. EFA and CFA were conducted to assess the dimensional structure of the QLQ-C30 and the results compared. The model of HRQOL tested using CFA was based on both the established structure of the QLQ-C30[15] and clinical considerations (described below).

Three items were excluded a priori from both the EFA and CFA. Item 28 (financial difficulties) was excluded from all analyses as it is neither a symptom nor a measure of functioning. The two global items (29 and 30) were also excluded because each item in the HSCS should represent a specific domain of HRQOL (functioning or symptom) rather than global quality of life.[3]

Exploratory approach

For the initial EFA, principal axis factoring (PAF) was used with a direct oblimin rotation to allow factors to be correlated. The suitability of the data for EFA was assessed using the Kaiser–Myer–Olkin measure of sampling adequacy and the Bartlett test of sphericity. Criteria for suitability are Kaiser–Myer–Olkin >0.8 and a *P*-value for Bartlett's χ^2 of less than 0.01.[16] Parallel analysis,[17] using the Monte Carlo PCA for Parallel Analysis software, was used to inform selection of factors. This involves computing mean eigenvalues from randomly generated sets of data (N=1,000) of the same size (number of items and number of observations) as the observed data set. Any factor obtained from the observed data set with an eigenvalue exceeding the corresponding eigenvalue generated from parallel analysis was considered for selection. A scree plot was also inspected. An item was considered to load on a factor if it had a pattern matrix loading greater than 0.3 and did not load on any other component.

We also conducted a sensitivity analysis involving all 15 combinations of: two extraction methods (PAF, maximum likelihood), and principal components analysis, and five rotation methods (oblimin, promax, varimax, equamax, and quartimax), comparing the degree of variability in solutions obtained due to variation in these technical parameters.

Confirmatory approach

A priori clinical considerations

The guiding principle here was to consider which aspects of functioning, symptoms, and side effects should be included in the HSCS, and hence the utility function of cancer-specific MAUI, in order for it to have face validity for economic evaluation of cancer treatments. Inclusion of dimensions was determined by three considerations: a) the dimensions available in the QLQ-C30; b) the patient's perspective (which symptoms, side effects, and aspects of functioning are considered important by patients in their overall assessment of quality of life); and c) the clinician's perspective (which dimensions matter when assessing the value of alternative treatments). Previous research has shown that patients[13] and clinicians[7] consider pain, fatigue, nausea/vomiting, constipation, and diarrhea to be important. All are available in the QLQ-C30. It is also well established that the various aspects of functioning are correlated with measures of overall quality of life.[14] Regression analysis has also revealed certain domains to be strong predictors of global quality of life, eg, emotional functioning and fatigue.[18,19]

The primary difference between clinical considerations using the confirmatory approach versus previous exploratory approaches is that in the confirmatory approach, they are incorporated a priori as part of the procedure to assess items for inclusion.

Established structure of the QLQ-C30

We defined the "conceptual model" as the arrangement of items on the QLQ-C30 into domains based on the established structure of the QLQ-C30[6] and the clinical considerations described above. We defined the "measurement model" as the subset of the conceptual model that was empirically tested using CFA.

The conceptual model to be used as a starting point for the QLQ-C30 was thus composed of the following eight latent variables and five single-item domains:

Functioning: physical functioning (items 1–5); role functioning (items 6 and 7); emotional functioning (items 21–24); social functioning (items 26–27); and cognitive functioning (items 20 and 25).

Symptoms: pain (items 9 and 19); fatigue (items 10, 12, and 18); nausea and vomiting (items 14 and 15); dyspnea (item 8); sleep (item 11); appetite (item 13); constipation (item 16); and diarrhea (item 17).

Items included a priori in the conceptual model and therefore excluded from measurement model: dyspnea, sleep, appetite, constipation, and diarrhea were considered of sufficient clinical importance for consideration in the HSCS, but as these domains are represented by single items (8, 11, 13, 16, and 17, respectively), these items were excluded from the measurement model.

CFA based on the conceptual models described above was conducted using the mean- and variance-adjusted weighted least squares estimation method (as recommended for ordinal data)[20] in Mplus Version 6. Correlations amongst the latent variables were not constrained, while correlations between error terms were fixed to 0. The fit of the model to the data was assessed using the following indices and their corresponding widely accepted guidelines indicating good model fit:[21] chi-squared statistic/degrees of freedom (less than 2); comparative fit index (>0.95); Tucker–Lewis index (>0.95); root mean square error of approximation (<0.05). If model fit was poor on any one of the measures, then factor loadings and residual correlations (those >0.1 considered noteworthy)[22] were examined in order to determine alterations to the model that improved fit. Modification indices were also examined to determine what other parameters might be estimated. The model was modified and retested until a

model was obtained that was conceptually meaningful and also adequately fitted the data.

Item assessment using Rasch analysis

Young et al[10] used a variety of techniques to select or reject items for the HSCS. These methods use Rasch analysis within dimensions identified by EFA. To address the aims of this paper, we conduct the Rasch analyses separately for the factor solutions obtained from EFA and CFA to further explore the consequence of these two approaches when applying Young et al's method to the QLQ-C30. These techniques are described in detail by Young et al[10] and interested readers are referred to step 2 of their guidance for deriving a MAUI. These are summarized briefly below.

In Rasch analysis, observed responses to items are assumed to reflect an underlying latent variable, such that the probability of endorsing an item is a monotonic increasing function of the underlying latent variable. Items that met the criteria described below were deemed to conform to the Rasch model[23] and were therefore retained for consideration in the HSCS.

All Rasch analyses were conducted using RUMM 2030[24] and were performed separately for the dimensions identified using EFA and CFA. All procedures and guidelines were consistent with those recommended by Pallant and Tennant.[25] The initial stage of Rasch analysis was conducted with the aim of determining whether any of the items exhibited problems with fit to the model, item response threshold ordering, or differential item functioning.[25] Local dependence was also assessed. Any items that exhibited such problems were considered for exclusion from the HSCS. See the Supplementary materials for further details regarding these criteria.

Results

Exploratory approach

Table 2 provides a summary of the results from the primary EFA (PAF extraction and oblimin rotation) and related

Table 2 Summary of item statistics based on the dimensions established using exploratory factor analysis

Item	Factors and loadings (exploratory factor analysis)[a]			Rasch			
				Location	Item fit	Differential item functioning[b]	Local dependency[c]
	Factor 1	Factor 2	Factor 3				
1	0.65	0.00	0.01	−0.92	0.88	Sex, site[d]	
2	0.75	−0.07	−0.07	−1.23	−1.34		3
3	0.82	−0.06	−0.15	0.96	0.06		2
4	0.65	0.03	0.14	0.14	0.73		10
5	0.52	−0.02	−0.05	3.40	0.95		
6	0.86	−0.08	0.02	−0.81	−3.89	Site	7
7	0.77	0.02	−0.02	−0.51	−1.37		6
8	0.15	0.17	0.08	Not included in Rasch analysis (weak factor loadings)			
9	0.49	0.06	0.05	Misfit			
10	0.53	0.26	0.25	−0.42	0.59		4
11	−0.02	0.34	0.02	Misfit			
12	0.25	0.38	0.34	−0.97	−0.37		18
13	0.01	0.12	0.62	−0.05	0.34		
14	−0.11	0.03	0.83	0.63	0.28	Site	15
15	−0.11	−0.01	0.77	1.64	−1.73		14
16	0.14	−0.03	0.18	Not included in Rasch analysis (weak factor loadings)			
17	0.02	−0.07	0.29	Not included in Rasch analysis (weak factor loadings)			
18	0.24	0.36	0.36	−1.26	−0.05		12
19	0.68	0.11	−0.02	−0.62	1.51		
20	0.15	0.41	0.29	Misfit			
21	−0.01	0.83	−0.10	0.28	0.29	Sex	
22	−0.01	0.91	−0.23	−0.28	−1.31	Sex, site	
23	−0.02	0.60	−0.01	Misfit			
24	−0.05	0.77	0.04	−0.001	1.49		
25	0.08	0.27	0.27	Not included in Rasch analysis (weak factor loadings)			
26	0.15	0.35	0.09	Misfit			
27	0.42	0.24	0.13	Misfit			

Notes: Rasch statistics are those obtained from the final analyses, ie, those with misfitting items removed. [a]Principal axis factoring extraction, direct oblimin rotation; [b]grouping variables exhibiting differential item functioning for the item are listed in this column; [c]values in this column represent numbers of items with which the item has a residual correlation following Rasch analysis; [d]cancer sites included prostate, breast, lung, and other.

Rasch analyses. The inter-item correlations were adequate for factor analysis (Kaiser–Myer–Olkin =0.892; Bartlett's χ^2=3,993.58, $P<0.0005$). Parallel analysis suggested the extraction of three factors, and this was supported by inspection of the scree plot. Items 8 (dyspnea), 16 (constipation), 17 (diarrhea), and 25 (memory) loaded weakly on all factors, while cross-loadings were observed for items 12 and 18 (both fatigue items).

The three factors identified for subsequent Rasch analysis were as follows:

- EFA Factor 1. Items 1–7, 9, 10, 19, and 27 (encompassing the physical and role functioning domains, the two pain items, one of the three fatigue items, and one of the two social functioning items);
- EFA Factor 2. Items 11, 20–24, and 26 (encompassing the emotional functioning domain, the insomnia item, one of the two cognitive functioning items, and one of the two social functioning items); and
- EFA Factor 3. Items 12–15 and 18 (encompassing two of the three fatigue items, the appetite loss item, and the two nausea/vomiting items). The two cross-loading items (fatigue 12 and 18) were assigned to this factor because they are symptoms that are more closely related to the items on this factor than Factor 2.

The results of EFA differed slightly depending on the extraction and rotation method used. Using all 15 combinations of methods: items 1–7, 9, and 19 loaded on Factor 1; items 11, 21–24, and 26 loaded on Factor 2; items 13–15 loaded on Factor 3; and items 8 and 16 exhibited weak loadings on all factors. There were a few noteworthy differences. Items 17 (diarrhea, Factor 3) and 25 (memory, Factor 2/Factor 3) had stronger loadings for PCA than for PAF and maximum likelihood, to the extent that, using a loading cutoff of 0.3, they would have been comfortably included in the PCA solution, but not PAF or maximum likelihood. For items 12 (weak) and 18 (tired), for all extraction methods loadings were strongest for Factors 2 and 3 except for when quartimax rotation was used; in this case, Factor 1 exhibited the dominant loadings. For items 10 (rest) and 27 (interfered with social activities), Factor 1 exhibited the dominant loading but strength of cross-loadings differed between extraction/rotation combinations, and the same for item 20 (concentration) except that Factor 2 dominated. Results are available from the authors on request.

Confirmatory approach

The factor loadings obtained from CFA are presented in Table 3. The loadings of all items on their respective factors were relatively strong and all statistically significant ($P<0.001$). Model fit was adequate (χ^2/df =2.79, comparative fit index =0.964, Tucker–Lewis index =0.953, root mean square error of approximation =0.075). Residual correlations and modification indices suggested additional relations between items 4 and 10, and items 2 and 3. Items 4 and 10 cover similar content (needing to rest), as do items 2 and 3 (trouble taking a long walk and short walk). Because items 4 and 10 were posited to load on different factors (Physical Functioning and Fatigue, respectively) cross-loadings were introduced for these items and domains, whereas because items 2 and 3 were posited to load on the same factor (Physical Functioning), the covariance between their error terms was estimated. Estimation of these cross-loadings and covariance resulted in improved model fit (χ^2/df=1.51, comparative fit index =0.990, Tucker–Lewis index =0.987, root mean square error of approximation =0.040).

The correlations between the eight factors are displayed in Table 4. Most noteworthy was the very high (0.86) correlation between role and physical functioning, suggesting that the items in these two factors may reflect a single factor.

Although the hypothesized eight-factor structure of the QLQ-C30 was generally supported, it was decided that the physical functioning domain (items 1–5) be combined with the role functioning domain (items 6 and 7) as well as item 10 for the purpose of Rasch analysis, based on the results above. Item 10 was not included in the fatigue domain (with items 12 and 18) for Rasch analysis. The other domains were subjected to Rasch analysis without any change from the factor specified a priori.

Rasch analysis

Based on EFA

The factor-level results of the Rasch analysis for the factors derived using EFA are shown in the left panel of Table 2. This table illustrates that Factors 1 and 2 required the removal of items to achieve adequate fit to the Rasch model. High residual correlations were observed between items 2 (long walk) and 3 (short walk), items 4 (stay in bed) and 10 (need to rest), items 6 (daily activities) and 7 (leisure activities), items 12 (weak) and 18 (tired), and items 14 (nausea) and 15 (vomiting). The correlations between items 6 and 7, items 12 and 18, and items 14 and 15 were unsurprising, as the traditional QLQ-C30 domain structure treats these as separate domains (role functioning, fatigue, and nausea/vomiting, respectively). The other two pairs of residual

Table 3 Summary of item statistics based on the dimensions established using CFA

Item	A priori factors, guided by conceptual model	CFA loadings	Rasch			
			Location	Item fit	Differential item functioning[a]	Local dependency[b]
1	Physical functioning	0.78	−1.01	0.79	Sex, site[c]	
2	Physical functioning	0.80	−1.33	−1.64		3
3	Physical functioning	0.79	0.90	−0.13		2
4	Physical functioning	0.58, 0.31[d]	0.06	0.71		10
5	Physical functioning	0.76	3.39	0.76		
6	Role functioning	0.94	−0.90	−3.62	Site	7
7	Role functioning	0.90	−0.59	−0.48		6
9	Pain	0.72	−0.38	1.42		
10	Fatigue	0.66, 0.39[d]	−0.52	0.97		4
12	Fatigue	0.87	0.22	−0.07	Site	
14	Nausea and vomiting	0.94	−1.20	−0.72	Site	
15	Nausea and vomiting	0.92	1.20	−0.21	Site	
18	Fatigue	0.88	−0.22	0.38		
19	Pain	0.97	0.38	0.24		
20	Cognitive functioning	0.89	−0.10	0.52		
21	Emotional functioning	0.89	0.28	0.29	Sex	
22	Emotional functioning	0.88	−0.28	−1.31	Sex, site	
23	Emotional functioning	0.67	Misfit			
24	Emotional functioning	0.86	−0.001	1.49		
25	Cognitive functioning	0.62	0.10	1.15		
26	Social functioning	0.63	0.323	1.03		
27	Social functioning	0.87	−0.323	0.71	Site	

Notes: The results are for the refined model, in which loadings for items 4 and 10 on both physical functioning and the covariance between items 2 and 3 were estimated. Rasch statistics are those obtained from the final analyses, ie, those with misfitting items removed. [a]Grouping variables exhibiting differential item functioning for the item are listed in this column; [b]values in this column represent numbers of items with which the item has a residual correlation following Rasch analysis; [c]cancer sites included prostate, breast, lung, and other; [d]estimate of loading on the non–a priori factor, ie, fatigue for item 4, physical functioning for item 10.
Abbreviation: CFA, confirmatory factor analysis.

correlations are also unsurprising, given the content of the items. No individual items exhibited misfit or disordered thresholds. Items 1, 6, 14, 21, and 22 exhibited differential item functioning (Table 2).

Based on CFA

Table 3 provides a summary of the results from the CFA and related Rasch analyses, and the factor-level results are shown in the right panel of Table 3. Only Factor 2 required the removal of items to achieve adequate fit to the Rasch model (see Table 5 for factor-level Rasch analysis statistics). High residual correlations were observed between items 2 (long walk) and 3 (short walk), items 4 (stay in bed) and 10 (need to rest), and items 6 (daily activities) and 7 (leisure activities). No individual items exhibited misfit or disordered thresholds. Items 1, 6, 12, 14, 15, 21, 22, and 27 exhibited differential item functioning (see Table 3).

Table 4 Correlations between factors obtained from the confirmatory factor analysis

	PF	RF	EF	SF	CF	Pain	Fatigue
RF	0.90						
EF	0.28	0.32					
SF	0.60	0.62	0.57				
CF	0.44	0.50	0.59	0.65			
Pain	0.73	0.78	0.39	0.58	0.43		
Fatigue	0.52	0.57	0.59	0.61	0.79	0.54	
NV	0.19	0.30	0.25	0.37	0.48	0.33	0.58

Abbreviations: CF, cognitive functioning; EF, emotional functioning; NV, nausea and vomiting; PF, physical functioning; RF, role functioning; SF, social functioning.

Discussion

The factor structures obtained from EFA and CFA followed by Rasch analysis were similar; however, CFA produced more readily interpretable solutions than EFA. Many of the discrepancies between the hypothesized factor structure in CFA and the clusters of items that emerged from EFA were eliminated when the factors obtained from EFA were subjected to Rasch analysis. For example, EFA Factor 2 originally comprised items 11, 20–24, and 26, but following Rasch analysis, this dimension was reduced to the emotional functioning domain of the QLQ-C30 (items 21–24). Item 23 was then further found to misfit and removed. The key point is that the confirmatory approach arrived at this solution more

Table 5 Summary of the factor-level statistics based on the dimensions established using exploratory (top panel) and confirmatory (bottom panel) factor analyses

Exploratory factor analysis	
Factor 1	Initial:
(1–7, 9, 10, 19, 27)	Item fit =2.44 (poor)
	Person fit =1.04 (good)
	Final (items 9, 27 removed):
	Item fit =1.71 (poor)[a]
	Person fit =0.92 (good)
Factor 2	Initial:
(11, 20–24, 26)	Item fit =2.82 (poor)
	Person fit =1.11 (good)
	Final (items 11, 20, 23, 26 removed):
	Item fit =1.41 (good)
	Person fit =1.34 (good)
Factor 3 (12–15, 18)	Initial:
	Item fit =0.85 (good)
	Person fit =0.79 (good)
Confirmatory factor analysis	
Factor 1 (1–7, 10)	Initial:
	Item fit =1.60 (poor)[a]
	Person fit =0.85 (good)
Factor 2 (21–24)	Initial:
	Item fit =2.28 (poor)
	Person fit =1.13 (good)
	Final (item 23 removed):
	Item fit =1.41 (good)
	Person fit =1.34 (good)
Factor 3 (26, 27)	Initial:
	Item fit =0.23 (good)
	Person fit =0.77 (good)
Factor 4 (20, 25)	Initial:
	Item fit =0.44 (good)
	Person fit =1.08 (good)
Factor 5 (9, 19)	Initial:
	Item fit =0.83 (good)
	Person fit =0.97 (good)
Factor 6 (12, 18)	Initial:
	Item fit =0.32 (good)
	Person fit =0.99 (good)
Factor 7 (14, 15)	Initial:
	Item fit =0.36 (good)
	Person fit =0.85 (good)

Notes: Item fit for both item and person represent the fit residual standard deviation, where a value greater than 1.5 is considered poor. [a]Although item fit was poor, no individual item exhibited misfit.

efficiently than the exploratory approach. Furthermore, the two adjustments to the measurement model tested in CFA that were required (namely, the estimation of the relations between items 4 and 10 and items 2 and 3) were readily identified and accommodated in the model.

The EFA results were found to differ somewhat depending on the method of extraction and rotation employed. Although these differences were not large, they may have had some impact on the item selection process. For example, the inclusion or exclusion of item 17 (diarrhea) and different decisions about which domain should include the fatigue items (12 and 18) may affect the composition of the HSCS.

Some aspects of the EFA solution were difficult to interpret. For example, the social functioning items loaded on different factors; specifically, item 26 (interfered with family life) loaded with physical/role functioning items and item 27 (interfered with social activities) loaded with emotional functioning items. Similarly, fatigue items loaded with nausea, vomiting, and lack of appetite. Although post hoc explanations of these relations are possible, and may well be causal (as discussed below), it is difficult to justify the inclusion of such items in the same domain for the purpose of selecting items for a utility instrument. For example, whether respondents experience interference with social activities is arguably a substantively different issue to whether respondents feel tense, and it seems inappropriate for these two items to be competing candidates for inclusion to represent the same factor in the HSCS. This means that judgment must be applied when using EFA as the factor analysis will establish "factors", and clinical input and interpretation is required to derive the "dimensions" from these factors. In contrast, in the CFA approach this guidance is provided at the outset to inform the factor analysis, meaning that the results directly represent the dimensionality of the measure. It is worth noting that three of the four items with weak EFA loadings (items 8, 16, and 17) were also three of the five items (along with items 11 and 13) that were excluded from the measurement model a priori.

EFA produced a solution that combined the physical (items 1–5) and role functioning domains (items 6 and 7) of the QLQ-C30. In the CFA, model fit was adequate with these two domains kept separate, although the two domains were very highly correlated. Residual PCA, as part of the Rasch analysis, confirmed that these are in fact two separate domains. One possible reason for this is that items 6 and 7 differ from items 1–5 in their "item difficulty", a phenomenon that would be more readily identified by Rasch analysis than factor analysis. An alternative explanation is that there exists a higher order factor that encompasses both physical and role functioning, or that there is some causal relation between these two factors. These latter possibilities are addressed further below, but are in any case more readily addressed using a confirmatory than an exploratory approach.

The confirmatory approach employed in the present analysis provided a structured role for clinical considerations and an explicitly articulated relation to the statistical and psychometric criteria used in the item selection process, whereas

in the previously employed exploratory approach, clinical considerations were less formally specified and explicitly integrated with the statistical analysis.

Rowen et al[9] in the derivation of EORTC 8D employed the input of a clinician to ensure the statistical results made sense clinically. In the present analysis, we have developed the structured integration of clinical considerations further into the predefined set of judgment criteria. Furthermore, by identifying certain items as of interest a priori allows a structured approach to the selection of items that are of clinical relevance but may not perform adequately in the statistical analysis. For example, although few respondents in this data set reported problems with diarrhea (item 17), the a priori inclusion of this item in the conceptual model allowed clinical considerations to override the statistical criteria. The importance of this is illustrated by the ALTTO trial, in which diarrhea was a critical side effect distinguishing trastuzimab from lapatinib.[19] The omission of diarrhea on statistical grounds, in this case, would result in the loss of potentially important information from the HSCS. This is not to say that the exploratory approach has little value in establishing the domain structure for a HSCS, particularly in cases where an instrument does not have a well-established domain structure.

Limitations

Our analysis was conducted on a sample of patients who were either Norwegian or Swedish, with two-thirds having primary cancer sites that were either breast or prostate and all having recurrent/metastatic cancer. Different results may be obtained from samples of patients with different profiles. Indeed, the EFA solution we obtained differed from that of Rowen et al,[9] who analyzed data from newly diagnosed multiple myeloma patients. Their factor solution may also have differed from ours for reasons related to analysis details, eg, use of parallel analysis to select the number of factors in the present case versus eigenvalues and variance explained. The conclusions drawn from the present analysis would be strengthened by replication using data from patients with a variety of cancer sites, stages, and treatments, and from various countries, using identical statistical techniques.

Conclusion

A confirmatory approach to determining dimensionality for the construction of a HSCS was found to be more efficient and to produce a more readily interpretable domain structure for the QLQ-C30. The confirmatory aspect of this prototype analysis will now be applied on a much larger scale as part of the Multi-Attribute Utility in Cancer (MAUCa) project, involving the pooling of a large number of international data sets covering a range of countries, cancer sites, and stages. Based on the results, a definitive HSCS will be determined. The results of the present analysis will guide this large-scale analysis only inasmuch as they support the use of the particular method – the specific composition of dimensions and psychometric properties of dimensions and items obtained will be assessed independently of the results of the present analysis. This will pave the way for valuation surveys that will provide country-specific utility weights for this HSCS, and thereby complete the provision of a preference-based measure derived from the QLQ-C30.

Acknowledgments

The Multi-Attribute Utility in Cancer (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, as outlined above: John Brazier, David Cella, Stein Kaasa, Georg Kemmler, Helen McTaggart-Cowan, Richard Norman, Stuart Peacock, Simon Pickard, Neil Scott, Martin Stockler, and Deborah Street. This research was supported by a National Health and Medical Research Council (NHMRC; Australia) Project Grant (632662). Monika Janda is supported by an NHMRC career development award 1045247. Professor King is supported by the Australian Government through Cancer Australia.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Blinman P, King M, Norman R, Viney R, Stockler M. Patients' preferences for cancer treatments: an overview of methods and applications in oncology. *Ann Oncol*. 2012;23(5):1104–1110.
2. The EuroQol Group. EuroQol – a new facility for the measurement of health related quality of life. *Health Policy*. 1990;16:199–208.
3. Brazier J, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. *Med Decis Making*. 2008; 28(1):113–126.
4. Brazier J, Usherwood T, Harper R, Thomas K. Deriving a preference-based single index from the UK SF-36 Health Survey. *J Clin Epidemiol*. 1998;51(11):1115–1128.
5. Young TA, Yang Y, Brazier JE, Tsuchiya A. The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. *Med Decis Making*. 2011;31(1):195–210.
6. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organisation for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365–376.
7. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. Oxford: Oxford University Press; 2005.

8. McTaggart-Cowan H, Teckle P, Peacock S. Mapping utilities from cancer-specific health-related quality of life instruments: a review of the literature. *Expert Rev Pharmacoecon Outcomes Res*. 2013;13(6):753–765.
9. Rowen D, Brazier J, Young T, et al. Deriving a preference-based measure for cancer using the EORTC-QLQC30. *Value Health*. 2011;14(5):721–731.
10. Young TA, Yang Y, Brazier JE, Tsuchiya A, Coyne K. The first stage of developing preference-based measures: constructing a health-state classification using Rasch analysis. *Qual Life Res*. 2009;18(2):253–265.
11. Cella D, Rosenbloom SK, Beaumont JL, et al. Development and validation of eleven symptom indexes to evaluate response to chemotherapy for advanced cancer. *J Natl Compr Canc Netw*. 2011;9(3):13–24.
12. Cella D, Paul D, Yount S, et al. What are the most important symptom targets when treating advanced cancer? A survey of providers in the National Comprehensive Cancer Network (NCCN). *Cancer Invest*. 2003;21(4):526–535.
13. Bjordal K, de Graeff A, Fayers PM, et al. 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. *Eur J Cancer*. 2000;36:1796–1807.
14. Kaasa S, Brenne E, Lund JA, et al. Prospective randomised multicenter trial on single fraction radiotherapy (8 Gy x1) versus multiple fractions (3 Gy x10) in the treatment of painful bone metastases. *Radiother Oncol*. 2006;29:278–284.
15. Gundy CM, Fayers PM, Groenvold M, et al. Comparing higher order models for the EORTC QLQ-C30. *Qual Life Res*. 2012;21(9):1607–1617.
16. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 5th ed. Boston: Allyn and Bacon; 2007.
17. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika*. 1965;30:179–185.
18. Ostlund U, Wennman-Larsen A, Gustavsson P, Wengstrom Y. What symptom and functional dimensions can be predictors for global ratings of overall quality of life in lung cancer patients? *Support Care Cancer*. 2007;15:1199–1205.
19. Tomasello G, de Azambuja E, Dinh P, Snoj N, Piccart-Gebhart M. Jumping higher: is it still possible? The ALTTO trial challenge. *Expert Rev Anticancer Ther*. 2008;8(12):1883–1890.
20. *Mplus User's Guide* [Computer Program]. Los Angeles, CA: Muthén and Muthén; 1998–2011.
21. Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36® Health Survey Manual and Interpretation Guide*. Boston: New England Medical Center, The Health Institute; 1993.
22. McDonald RP. *Test Theory: A Unified Treatment*. New Jersey: Lawrence Erlbaum; 1999.
23. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press; 1960.
24. *RUMM 2020* [Computer Program]. Perth: RUMM Laboratory; 2003.
25. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol*. 2007;46:1–18.
26. Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparisons in valuing EQ-5D health states: a review and analysis. *Value Health*. 2009;12(8):1194–1200.

Supplementary material

Rasch analysis criteria

Poor item fit

The overall fit of the Rasch model was examined using the item–trait interaction χ^2 statistic. Good model fit was indicated by a nonsignificant chi-squared statistic. A Bonferroni correction was applied to the criterion of significance with the alpha value (0.05) divided by the number of items. The presence of misfitting items or persons was indicated by a fit residual standard deviation value of 1.5 or above. Items with individual Fit Residual values exceeding 2.5 were removed from the Rasch analysis. Persons with fit residuals that exceeded 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 of the model was nonsignificant. Any items excluded due to misfit were kept aside and assessed according to other criteria, including descriptive statistics and clinical considerations (described to follow).

Assessment of response format

An appropriately functioning item requires a response format that respondents use in a consistent manner. Examining response thresholds – the points at which each consecutive response category for an item is equally likely to be endorsed – allows the assessment of response format in this regard. For an appropriately functioning item, the response thresholds between successive categories should be ordered, such that the threshold between categories 1 and 2 falls below the threshold between categories 2 and 3, and so on. A disordered response threshold indicates that respondents are not selecting response categories expected according to their overall scale score.

Invariance of item functioning across different groups

For an item to be included in the HSCS, the probability of selecting a certain response category for a given value of the latent trait should be invariant across groups. If it is not, the item exhibits differential item function (DIF). DIF is 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. If two or more groups showed a consistent difference in item responses across the range of values for the latent variable, this is known as "uniform DIF". "Non-uniform DIF" occurs when the differences between groups vary over the range of values of the latent variable. In RUMM 2020, DIF is assessed using two-way analysis of variance, with predicted score compared across the different levels of the grouping variable and across different levels of the latent trait (where individuals are grouped into a number of "class intervals" based on their latent trait score [35]). The data were examined for DIF across sex and cancer site. (DIF across country is an important issue but has been examined previously.) Because cross-population comparisons using the HSCS are desirable, any items exhibiting DIF were excluded from the HSCS.

Local dependence

Local dependence among items, indicating an association above and beyond that shared by the underlying trait, was assessed by inspection of the residual correlation matrix for values exceeding 0.3.