

Developing the Breast Utility Instrument to Measure Health-Related Quality-of-Life Preferences in Patients with Breast Cancer: Selecting the Item for Each Dimension

Teresa C. O. Tsui¹, Maureen E. Trudeau, Nicholas Mitsakakis¹, Murray D. Krahn and Aileen M. Davis

Abstract

Introduction. Generic preference-based instruments inadequately measure breast cancer (BrC) health-related quality-of-life preferences given advances in therapy. Our overall purpose is to develop the Breast Utility Instrument (BUI), a BrC-specific preference-based instrument. This study describes the selection of the BUI items. **Methods.** A total of 408 patients from diverse BrC health states completed the EORTC QLQ-C30 and BR45 (breast module). For each of 10 dimensions previously assessed with confirmatory factor analysis, we evaluated data fit to the Rasch model based on global model and item fit, including threshold ordering, item residuals, infit and outfit, differential item functioning (age), and unidimensionality. Misfitting items were removed iteratively, and the model fit was reassessed. From items fitting the Rasch model, we selected 1 item per dimension based on high patient- and clinician-rated item importance, breadth of item thresholds, and clinical relevance. **Results.** Global model fit was good in 7 and borderline in 3 dimensions. Separation index was acceptable in 4 dimensions. Item selection criteria were maximized for the following items: 1) physical functioning (trouble taking a long walk), 2) emotional functioning (worry), 3) social functioning (interfering with social activities), 4) pain (having pain), 5) fatigue (tired), 6) body image (dissatisfied with your body), 7) systemic therapy side effects (hair loss), 8) sexual functioning (interest in sex), 9) breast symptoms (oversensitive breast), and 10) endocrine therapy symptoms (problems with your joints). **Conclusions.** We propose 10 items for the BUI. Our next steps include assessing the measurement properties prior to eliciting preference weights of the BUI.

Highlights

- A previous confirmatory factor analysis established 10 dimensions of the European Organisation for Research and Treatment of Cancer (EORTC) core quality of life questionnaire (QLQ-C30) and its breast module (BR45).
- In this study, we selected 1 item per dimension based on fit to the Rasch model, patient- and clinician-rated item importance, breadth of item thresholds, and clinical relevance.
- These items form the core of the future Breast Utility Instrument (BUI).
- The future BUI will be a novel breast cancer-specific preference-based instrument that potentially will better reflect women's preferences in clinical decision making and cost utility analyses.

Corresponding Author

Teresa C. O. Tsui, Hospital for Sick Children, Peter Gilgan Centre for Research and Learning, 686 Bay Street, 11th Floor, Toronto, ON M5G 0A4, USA; (teresa.tsui@utoronto.ca).



Keywords

breast cancer, health-related quality of life, condition-specific preference-based instrument, Rasch analysis, item selection

Date received: January 18, 2022; accepted: October 20, 2022

The EORTC QLQ-C30 (QLQ-C30) is an established general cancer health-related quality-of-life (HRQoL) questionnaire.¹ In patients with breast cancer (BrC), the QLQ-C30 and EORTC breast module BR45² are used together to describe general cancer and BrC-specific functional and symptomatic impairments resulting from the disease and/or its treatments in clinical trials. There is currently no satisfactory method of using EORTC QLQ-C30 and BR45 questionnaire responses to inform drug reimbursement decisions.

Decisions regarding the public reimbursement of drugs rely largely on the incremental cost-effectiveness ratio of an economic evaluation, the additional cost of a new treatment relative to the quality-adjusted life-years (QALYs) gained. QALYs are health utilities multiplied by time, a key measure of effectiveness in cost-utility analyses. One definition of utility is a preference-based measure of HRQoL, anchored at 0 and 1 (dead and full health). Utility instruments include valuations for health outcomes. The validity and outcomes of cost-utility analyses are sensitive to the utility value,^{3,4} particularly for preference-sensitive decisions, when evidence for alternative outcomes is conflicting or weak.⁵⁻⁷

Generic preference-based instruments such as the EuroQol-5-dimension (EQ-5D)^{8,9} are widely used yet may be less sensitive at detecting condition-specific changes in milder health states^{10,11} and also lack condition-specific content. To address this limitation, condition-specific preference-based instruments (CSPBIs) have been developed, commonly derived from existing psychometric instruments.¹²

To create a CSPBI, a reduced health state classification system composed of representative items amenable to valuation is often a precursor step.¹³ Reduced form instruments whether preference based or psychometric are also in high demand as they improve feasibility,¹⁴ whether completed in clinical settings routinely or part of clinical trials. At least 3 preference-based instruments have been derived from the general cancer QLQ-C30,^{15,16} which are used to measure utilities in cost-utility analyses. However, these instruments are limited as they lack content from symptoms specific to patients with BrC reflected in the BR45 module.²

A CSPBI is important in BrC because its treatments are preference sensitive. These include 1) the choice between mastectomy and breast-conserving surgery, which is related to how a woman values the preservation of the breast¹⁷; 2) timing of reconstruction, where immediate reconstruction postmastectomy shows better psychological outcomes but at a higher complication rate^{18,19}; and 3) the benefits of adjuvant therapy in modestly reducing recurrence rate needs to be weighed against hormonal and fertility effects of treatment.²⁰

There is also strong evidence that decision and cost-effectiveness analyses are often sensitive to utilities for health outcomes in BrC. Breast reconstruction decisions are sensitive to utilities for reconstruction outcomes,²¹ and treatments for early and advanced disease are sensitive to utilities for health outcomes.^{22,23} As such, decisions about BrC treatments are preference sensitive. This applies across the spectrum of BrC and at individual and health care system levels.

We previously expanded the conceptual framework of Brazier et al.¹³ to derive a preference-based instrument. Our framework has 4 phases of instrument development: I) develop initial questionnaire items, II) establish

Toronto Health Economics and Technology Assessment (THETA) Collaborative, University Health Network, Toronto, Ontario, Canada (TCOT, MDK); Canadian Centre for Applied Research in Cancer Control (TCOT); Hospital for Sick Children, Toronto, Ontario, Canada (TCOT); Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada (MT); Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada (NM); Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada (NM); Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada (AMD). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided in part by Dr. Kathleen Pritchard, CM, MD, FRCPC, the Toronto Health Economics and Technology Assessment Collaborative (THETA) Fund for Excellence (No. 5790 6839 0706), Toronto General Hospital Research Institute, and support from the Sunnybrook Research Institute. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. AMD and MDK are joint senior authors.

dimension structure, III) reduce items per dimension, and IV) value and model health state utilities.²⁴ Following this framework, Rasch analysis and other psychometric criteria are applied in phase III to reduce items per dimension after establishing a dimension structure.

Our overall objective is to develop the Breast Utility Instrument (BUI), a BrC-specific preference-based instrument. This will be the first BrC-specific preference-based instrument derived from the EORTC QLQ-C30 and BR45. This would allow responses from the EORTC QLQ-C30 and BR45 to be derived as a utility score in the future. In this study, we describe selecting 1 item representing each dimension using the same sample of our previous confirmatory factor analysis of the QLQ-C30 and BR45.²⁵

Methods

Overview

We evaluated the fit of item responses from the QLQ-C30 and BR45 to the Rasch model. For items that fit the Rasch model, we then selected 1 item per dimension based on patient- and clinician-rated item importance,²⁶ the range of item thresholds, and clinical relevance.

Parent Instruments

We started with patient responses to the QLQ-C30 version 3,¹ a 30-item general cancer HRQoL patient-reported instrument with functioning and symptom subscales, and global health items. The BR45 is a BrC-specific module updated² from the BR23²⁷ to reflect current treatments.

Data. A cross-sectional convenience sample of women 18 y and older with invasive BrC attending medical oncologists' outpatient clinics at the Sunnybrook Odette Cancer Centre were enrolled to complete paper questionnaires of QLQ C30 and the BR45. Patients were stratified in 1 of 5 mutually exclusive health states: 1) first year after primary BrC diagnosis treated with curative intent (I), second to fifth year after primary BrC diagnosis (II–V), sixth year onward after primary BrC diagnosis (VI), metastatic BrC (M), and first year after recurrence of BrC or new BrC.^{28,29} Patients had no other primary cancer within the past 5 y and understood English or had a translator.

A random subset of patients with BrC ($n = 81$) and clinicians working with patients with BrC ($n = 13$) rated the importance of all items in EORTC QLQ-C30 and

BR45 on a 5-point scale (0 = *not applicable*, 5 = *very important*; Appendix A). Clinicians were asked to rate the importance of items as relevant to patients' experiences. Clinicians completed item importance ratings on a secure web form. No demographic characteristics of clinicians were collected to protect their anonymity.

The mean importance of items rated by patients and clinicians were converted to rankings (1 = most important) to improve interpretability.^{26,30}

This study was approved by research ethics boards at the University Health Network, Sunnybrook Health Sciences Centre, and the University of Toronto. All participants provided informed consent.

Ten dimensions of the BUI. The 10 dimensions of the BUI derived from the QLQ-C30 and BR45 include functioning dimensions (physical and role, emotional, social, body image, sexual functioning and enjoyment) and symptom dimensions (fatigue, pain, systemic therapy side effects, arm and breast, endocrine therapy; Table 1). The response options were maintained from the original QLQ-C30 and BR45, scored 1 = *not at all* to 4 = *very much*. High scores for a functioning scale represent high functioning; high scores for a symptom scale represent a high level of symptoms or problems.^{2,27}

Analyses. We used a priori criteria to select items. We first conducted Rasch analysis as recommended by Tennant and Conaghan³¹ and applied by developers of CSPBIs.^{15,16,32–62} Data that fit the Rasch model satisfy the condition for interval-level data.³¹ In addition, item and person parameters are independent of the sample.⁶³ We used the partial credit model for multiple response options.⁶⁴

For each of the ten dimensions, we iteratively examined overall model and item fit as follows:

1. Ordering of response thresholds. We examined item characteristic curves and item threshold maps to visualize the difficulty and hierarchy of items. Item response categories were examined to assess if they produced sequentially ordered item thresholds (i.e., the point at which there is a 50% probability of the higher or lower response). Disordered response categories were collapsed and recoded. These items were retained and considered for selection.
2. Rasch model fit. We evaluated the χ^2 goodness-of-fit statistic with 100 simulated values and Bonferroni-adjusted significance levels. The critical P value was 0.05/(number of items per dimension). Given the

Table 1 Patient and Clinician Mean Importance Ratings of the EORTC QLQ-C30 and BR45 Items^{a,b}

Dimension and Item Topic	Patients' Mean Item Importance (<i>n</i> = 81)	Clinicians' Mean Item Importance (<i>n</i> = 13)	Welch <i>t</i> Test <i>P</i> Value
Physical and role functioning			
PF1. Trouble doing strenuous activities.	3.79	3.58	0.413
PF2. Trouble taking a long walk.	3.86	3.85	0.248
PF3. Trouble taking a short walk.	3.14	4.08	0.003 ^c
PF4. Need to stay in bed or a chair during the day.	3.67	4.33	0.118
PF5. Need help with eating, dressing, washing yourself or using the toilet.	0	4.42	Insufficient sample size for comparison
RF6. Limited in doing either your work or other daily activities.	3.38	4.08	
RF7. Limited in pursuing your hobbies or other leisure time activities.	3.50	3.92	0.122
FA10. Need to rest.	3.32	3.55	0.342
Emotional functioning			
During the past week:			
EF21. Did you feel tense?	3.38	3.67	0.243
EF22. Did you worry?	3.33	4.23	<0.001 ^c
EF23. Did you feel irritable?	3.33	3.85	0.036 ^c
EF24. Did you feel depressed?	3.61	4.62	<0.001 ^c
ET55. Have you had mood swings?	3.38	3.75	0.153
Social functioning			
During the past week:			
SF26. Physical condition or medical treatment interfered with your family life.	3.50	4.17	0.083 ^c
SF27. Physical condition or medical treatment interfered with your social activities.	3.67	3.77	0.72
Body image			
BI39. Felt physically less attractive as a result of your disease or treatment.	3.39	3.92	0.079
BI40. Felt less feminine as a result of your disease or treatment.	3.41	3.62	0.481
BI41. Problems looking at yourself naked.	3.63	3.75	0.676
BI42. Dissatisfied with your body.	3.63	3.67	0.895
Sexual functioning and enjoyment			
During the past week:			
SX44. Interested in sex.	4.20	3.73	0.107
SX45. Sexually active (with or without intercourse).	4.24	3.45	0.005 ^c
SE46. Sex been enjoyable.	4.03	3.73	0.236
Symptom dimensions			
Fatigue			
During the past week:			
FA12. Felt weak?	3.17	4.15	<0.001 ^c
FA18. Have you felt tired?	3.35	3.92	0.032 ^c
ET 56. Have you felt dizzy?	3.40	3.90	0.168
Pain			
During the past week:			
PA9. Did you have pain.	3.56	4.15	0.043 ^c
PA19. Pain interfered with your daily activities.	3.47	4.38	0.002 ^c
Systemic therapy side effects			
SYS31. Dry mouth.	3.43	3.64	0.488
SYS32. Food and drink tasted different than usual.	3.67	3.58	0.841
SYS33. Eyes been painful, irritated or watery.	3.25	3.50	0.299
SYS34. Lost any hair.	4.00	3.92	0.779
SYS36. Felt ill or unwell.	3.14	4.00	0.004 ^c
SYS38. Headaches.	3.50	3.69	0.536

(continued)

Table 1 (continued)

Dimension and Item Topic	Patients' Mean Item Importance (n = 81)	Clinicians' Mean Item Importance (n = 13)	Welch t Test P Value
Arm and breast symptoms			
ARM47. Pain in your arm or shoulder.	3.38	3.54	0.502
ARM48. Swollen arm or hand.	3.11	3.92	0.003 ^c
ARM49. Problems raising your arm or moving it sideways.	3.17	4.00	0.007 ^c
BR50. Pain in the area of your affected breast.	3.14	3.50	0.157
BR51. Area of your affected breast been swollen.	3.50	3.82	0.390
BR52. Area of your affected breast been oversensitive.	3.43	3.90	0.148
BR53. Skin problems on or in the area of your affected breast (e.g., itchy, dry, flaky).	3.33	3.90	0.094
Endocrine therapy symptoms			
SYS37. Hot flushes.	3.94	3.92	0.938
ET54. Sweated excessively.	3.85	3.83	0.971
ET63. Problems with your joints.	3.74	4.00	0.382
ET64. Stiffness in your joints.	3.68	4.00	0.244
ET65. Pain in your joints.	3.75	4.17	0.125
ET66. Aches or pains in your bones.	3.64	4.08	0.183
ET67. Aches or pains in your muscles.	3.71	4.00	0.310
ET68. Gained weight.	3.71	3.67	0.859
ET69. Weight gain been a problem for you.	3.88	4.08	0.441

^aPatients rated the importance of items on a 0 to 5 scale (0 = *not applicable*, 5 = *very important*).

^bClinicians rated item importance on the same 0 to 5 scale as relevant to patients' experiences.

^cSignificant at $\alpha = 0.05$.

relatively large sample size, the χ^2 statistic can overestimate lack of model fit⁶⁵; therefore, we also considered the person separation index (PSI), where >0.7 was considered acceptable fit.³¹

- Item fit statistics. We examined item residuals and infit and outfit statistics, where infit is the weighted mean of the standardized squared residuals and outfit is the mean of the standardized squared residuals.⁶⁶ Misfitted items had large residuals >2.5 or <-2.5 or infit/outfit statistics beyond acceptable values, that is, <0.7 or >1.3 .³¹
- Differential item functioning (DIF). Item bias or DIF occurs when individuals with the same level of HRQoL systematically respond differently based on specific characteristics.³¹ DIF dependent on age (≤ 50 y and >50 y) was evaluated by specifying logistic regression equations to predict item responses from person parameters and the age group variable. Proportional odds models were compared using likelihood ratios.⁶⁷
- Unidimensionality. We performed principal components analysis (PCA) of the item residuals, expecting that no meaningful factors would remain in the residuals after fitting the data to the Rasch model. The

t-statistics of the person scores of positive (>0.30) and negative (<-0.30) factor loadings were also compared to determine the percentage significant at the 0.05 level. Less than 5% of t-statistics >1.96 or <-1.96 indicate strict unidimensionality.^{31,68}

Items considered for deletion (misfitted items, DIF) were iteratively removed one at a time, and overall model fit and item fit were reevaluated. We retained both items for pain and social functioning since 2 items is the minimum number for a dimension.⁶⁹

We used R version 4.1.0 for the analysis,⁷⁰ with packages psych,⁷¹ eRm,⁷² ltm,⁷³ and lordif.⁶⁷

Item selection criteria included the following:

- range of item thresholds, where a wider range was a better representation of construct severity;
- item goodness of fit, particularly infit 0.7 to 1.3;
- patient-rated item importance and clinician-rated item importance (items ranked most important by patients were prioritized over clinician rankings); and
- psychometric criteria:
 - Absence of floor and ceiling effects. Floor effects occur when a large proportion of respondents

select the worst possible score (e.g., “very much” difficulty in performing physical activity). Ceiling effects occur when a large proportion of respondents select the best possible score (e.g., “not at all” for pain), and

- b. Correlation of item to dimension (>0.70).
5. Items meeting the *a priori* criteria were reviewed by the multidisciplinary research team with expertise in patient outcome measurement, biostatistics, health economics, general internal medicine, and breast medical oncology. The clinical expert (MT) provided input on clinical relevance, particularly when there was no clear single item to represent a dimension.

Sample size. For the Rasch analysis, the sample size of more than 400 participants allows for very stable item calibration, where the width of the 99% confidence interval is smaller than 1 logit.⁷⁴ For patient and clinician item importance, we aimed to recruit 50 patients and 10 experts^{30,75,76} to provide a spectrum of importance ratings.

Results

Sample Characteristics

Table 2 outlines the details of all patients ($N = 408$) and the subset of patients who completed item importance ($n = 81$). Overall, they represent women with BrC who were well educated and predominantly on adjuvant endocrine therapy. The mean (s) age was 59.1 (11.6) y, with 80% completing at least college education. Most patients were diagnosed in pathological stage 1A (37.0%) or IIA (25.5%), which is comparable to the incidence of BrC stages in Ontario, Canada,^{77,78} with 64.2% on a form of adjuvant therapy and 57% on endocrine therapy. Most patients were in their second to fifth year postdiagnosis (31.1%). A higher proportion of patients in the sample had metastatic disease than the patients in a development study of the BrC health states (25.2% v. 19.4%).²⁸ Appendix B lists the specific therapies that patients were taking.

The subset of patients ($n = 81$) who rated item importance were of comparable age, biomarker status, and comorbidity status as all participants (Table 2). The item-importance sample consisted of a smaller percentage with a graduate or professional degree than the full sample (29.6% v. 37.3%), fewer in the metastatic health state (16.0% v. 25.2%), and a larger percentage diagnosed with BrC from 5 to 9 y (34.6% v. 22.3%). Appendix C

shows the mean item importance and rank importance for patients and clinicians.

The 13 clinicians who completed importance ratings were 5 medical oncologists, 1 radiation oncologist, 1 surgical oncologist, 2 medical oncology fellows, 2 nurses, 1 physician assistant, and 1 social worker, predominantly representing the medical oncology clinical staff.

Patient and Clinician Item Importance ratings

Overall, within each dimension, patients and clinicians had comparable item importance ratings based on Welch t tests. When ratings differed (e.g., pain dimension), clinicians rated item importance as significantly higher than patients did (Table 1). The highest rated items for patients were on sexual functioning and enjoyment and hair loss. For clinicians, the highest rated items were needing help with eating, dressing, etc.; feeling depressed; and pain interfering with daily activities. Across all items, 30% of patients rated items as 0 (not applicable), 1 (slight), or 2 (mild); therefore, these scores were removed prior to calculating the mean item importance of responses rated 3 (moderate), 4 (strong), and 5 (severe) of all 81 patients to place more emphasis on items with greater impairment on HRQoL.³⁰

Overview Fit to Rasch Model

Appendix D and Table 3 show the initial and final Rasch models and fit statistics for each dimension, respectively. Appendix E shows the person-item map and histogram of person location estimates by dimension. Appendix F shows the item characteristic curves of items in each dimension. Overall, the global Rasch model fit was good in 7 and borderline in 3 dimensions. Person separation reliability was acceptable in 4 dimensions after item removal and response-level collapsing. DIF by age (≤ 50 , >50) was present in 8 items, described within their dimensions below and in Appendix G. The magnitude of DIF items was all negligible, with McFadden's $R^2 < 0.13$.⁸⁰

Item Selection for Each Dimension

Table 4 summarizes the results of mean item importance rankings and psychometric and Rasch criteria used to select 1 item per dimension. Subsequently, we describe the results for each dimension.

Physical and role functioning dimension. The initial 8-item model had good global fit to the Rasch model (χ^2

Table 2 Participant Characteristics and Comparator Population-Level Characteristics

	All Patients (<i>N</i> = 408)	Subset of All Patients, Item Importance (<i>n</i> = 81)	Population Comparators ^a	Population, Reference
Age, y				
\bar{x} (<i>s</i>)	59.1 (11.6)	60.12 (11.1)	61.5 (13.8)	Ontario women with breast cancer, Seung et al. ⁷⁷
Range	25–93			
Marital status, <i>n</i> (%)			<i>N</i> = 5,812,755	Canadian census
Single	50 (12.3)	12 (14.8)	1,493,605 ^b (25.7)	2016, women >15 y in Ontario 100% of data ⁷⁹
Married or common-law	262 (64.2)	48 (59.3)	3,218,800 (55.4)	
Divorced	36 (8.8)	7 (8.6)	400,935 (6.9)	
Separated	16 (3.9)	2 (2.5)	190,535 (3.3)	
Widowed	31 (7.6)	8 (9.9)	508,880 (8.8)	
Missing/did not answer	13 (3.2)	4 (4.9)	0 (0)	
Highest level of education, <i>n</i> (%)			<i>N</i> = 5,695,685	Canadian census
Elementary school	3 (0.7)	1 (1.2)	973,670 (17.1)	2016, women >15 y in Ontario, 25% sampled data ⁷⁹
High school	53 (13.0)	14 (17.3)	1,540,770 (27.1)	
Trade or apprentice	8 (2.0)	1 (1.2)	193,120 (3.4)	
College or undergraduate university	174 (42.6)	37 (45.7)	2,614,965 (45.9)	
Graduate or professional degree	152 (37.3)	24 (29.6)	373,160 (6.6)	
Missing/did not answer	18 (4.4)	4 (4.9)	0 (0)	
Years since first BrC diagnosis, <i>n</i> (%)				Women with breast cancer, Lidgren et al. ²⁸
< 5	229 (56.1)	35 (43.2)	183 (53.0)	
5 to 9	91 (22.3)	28 (34.6)	88 (25.5)	
10 to 14	38 (9.3)	10 (12.3)	74 (21.4)	
15 to 19	19 (4.7)	3 (3.7)		
20 to 25	19 (4.7)	3 (3.7)		
25 +	4 (1.0)	0 (0.0)		
Missing ^c	8 (2.0)	2 (2.5)		
Charlson comorbidity index (CCI), <i>n</i> (%)				Ontario women with breast cancer, Seung et al. ⁷⁷
0	366 (89.7)	71 (87.7)	\bar{x} (<i>s</i>): 0.64 (1.2)	
1	25 (6.1)	7 (8.6)		
2	8 (2.0)	1 (1.2)		
3	2 (0.5)	0 (0.0)		
Missing	7 (1.7)	2 (2.5)		
\bar{x} (<i>s</i>)	0.12 (0.4)	0.11 (0.4)		
Health state, <i>n</i> (%) ^d				Women with breast cancer, development of health states Lidgren et al. ²⁸
I	81 (19.9)	13 (16.0)	72 (20.9)	
R	9 (2.2)	2 (2.5)	21 (6.1)	
II–V	127 (31.1)	29 (35.8)	185 (53.6)	
VI +	88 (21.6)	24 (29.6)		
M	103 (25.2)	13 (16.0)	67 (19.4)	
Menstrual status, <i>n</i> (%)				Women with breast cancer, development of BR45, Bjelic-Radisic et al. ²
Premenopausal	40 (9.8)	9 (11.1)	59 (23.6)	
Postmenopausal	309 (75.7)	58 (71.6)	178 ^e (71.2)	
I don't know	32 (7.8)	7 (8.6)	11 (4.4)	
Missing	27 (6.6)	7 (8.6)	2 (0.8)	
Pathological stage at initial surgery, <i>n</i> (%)				Ontario women with breast cancer, Seung et al. ⁷⁷
IA	151 (37.0)	20 (28.6)	13,989 (40.7)	
IB	4 (1.0)			
IIA	104 (25.5)	28 (40.0)	12,819 (37.3)	
IIB	69 (16.9)	10 (14.3)		
IIIA	45 (11.0)	7 (10.0)	4,508 (13.1)	
IIIB	5 (1.2)	0 (0)		
IIIC	15 (3.7)	4 (5.7)		
IV	12 (2.9)	1 (1.4)	1673 (4.9)	
No surgery/unknown	3 (0.7)		1265 (3.7)	

(continued)

Table 2 (continued)

Biomarkers, <i>n</i> (%)	All Patients (<i>N</i> = 408)	Subset of All Patients, Item Importance (<i>n</i> = 81)	Population Comparators ^a	Population, Reference
	Positive	Negative	Disease subtype	
Estrogen receptor	329 (47.8)	64 (48.1)	Hormone + ,	Ontario women with breast cancer, Seung et al. ⁷⁷
Progesterone receptor	294 (42.7)	57 (42.9)	HER2-: 22,247 (64.8)	
HER-2 receptor ^f	66 (9.6)	12 (9.0)	HER2 + : 4902 (14.3)	
			Triple negative: 3277 (9.5)	
			Unknown subtype: 3914 (11.4)	
Surgery, <i>n</i> (%) ^g				Women with breast cancer, development of BR45, Bjelic- Radisic et al. ²
Breast-conserving surgery	281 (53.8)	63 (54.8)	117 ^h (40.5)	
Mastectomy	196 (37.5)	35 (30.4)	96 ⁱ (33.2)	
Mastectomy: prophylactic	26 (5.0)	7 (6.1)		
Other	15 (2.9)	10 (8.7)	37 (12.8)	
Missing	4 (0.8)		39 (13.5)	
Surgery-axillary, <i>n</i> (%)				Women with breast cancer, development of BR45, Bjelic- Radisic et al. ²
Axillary lymph node dissection (ALND)	125 (30.6)	23 (28.0)	92 (50.3)	
Sentinel lymph node biopsy (SLNB)	240 (58.8)	54 (65.9)	91 (49.7)	
SLNB and ALND	11 (2.7)	5 (6.1)		
Intent of systemic therapy, <i>n</i> (%)				
Neoadjuvant	26 (6.4)	4 (4.9)		
Adjuvant	262 (64.2)	58 (71.6)		
Palliative	92 (22.5)	12 (14.8)		
No treatment-active surveillance	28 (6.9)	7 (8.6)		
Regimen, <i>n</i> (%)				
Chemotherapy	96 (17.1)	11 (10.6)		
Endocrine therapy	321 (57.0)	66 (63.5)		
No treatment-active surveillance	28 (5.0)	6 (5.8)		
Radiotherapy	17 (3.0)	4 (3.8)		
Targeted therapy	94 (16.7)	14 (13.5)		
Zoledronic acid only	7 (1.2)	3 (2.9)		

^aPopulation comparators were mostly from women with breast cancer, except that marital status and highest level of education comparators were drawn from the 2016 Canadian Census.

^bFrom 2016 Canadian census data: never married.

^cReferral from another center. Date and month was approximate.

^dMutually exclusive health states: 1) first year after primary BrC diagnosis treated with curative intent (I), second to fifth year after primary BrC diagnosis (II-V), sixth year onward (VI), metastatic diseases (M), recurrence of BrC (R).

^eFive breast tumors were HER-2 equivocal.

^fFour hundred five people had a combined 522 surgeries. Three people did not receive surgery.

^gOne hundred fifty postmenopausal and 28 treatment-related menopause, for a total of 178.

^hOne hundred four breast-conserving surgeries and 13 oncoplastic breast-conserving surgeries, for a total of 117.

ⁱForty-nine simple mastectomies and 47 mastectomies and reconstruction surgeries, for a total of 96.

$P = 0.04$) and reliability (PSI = 0.811). All items had small mean standardized fit residuals (-0.019) and good infit. PF3 (*Trouble taking a short walk*) had low outfit (0.497) and ceiling effects (84.8%). PF5 (*Need help*

eating, dressing etc.) had low outfit (0.275) and ceiling effects (96.6%). Two items had uniform DIF by age: PF2 (*Trouble taking a long walk*) and PF7 (*Limited in pursuing hobbies*), although the magnitude of DIF was

Table 3 Summary of Final Fit of the Data to the Rasch Model^a

Dimension	No. of Items	Goodness of Fit		Person Separation Index		Items		Persons		Residuals		Misfit Persons, %	Strict Unidimensionality, %; <i>t</i> stats >1.96 or <-1.96
		χ^2	P Value	Bonferroni-Corrected P Value	PSI	Location		Location		Standardized Item Fit Residual			
						\bar{x}	<i>s</i>	\bar{x}	<i>s</i>	\bar{x}	<i>s</i>		
PRF no q7	7	7804	0.29	0.007	0.776	0	1.564	-2.160	2.110	-0.019	0.836	2.07	0.40
EF no q55	4	1727.49	0.01	0.013	0.736	0	0.609	-1.416	2.441	-0.006	0.866	2.14	0.85
SF	2	215.74	0.01 ^a	0.025	0.603 ^b	0	0.477	-2.044	3.218	0.000	0.686	0.49	N/A
FA	3	323.91	0.02	0.017	0.526 ^b	0	1.501	-1.620	2.404	-0.015	0.788	2.33	N/A
PA	2	108.37	0.01 ^b	0.025	0.551 ^b	0	0.610	-1.947	3.042	-0.010	0.714	0.83	N/A
BI	4	657.26	0.01	0.013	0.767	0	0.393	-0.786	2.548	-0.007	0.890	2.97	0.50
SXE	3	445.25	0.01 ^b	0.017	0.543 ^b	0	0.661	-1.584	2.015	0.001	0.853	2.69	N/A
SYS-all items recorded no q32, 34, 36, 38	6	1097	0.01	0.008	0.392 ^b	0	0.538	-1.218	1.220	-0.012	0.924	2.08	0
ARMBR	7	22873.37	0.06	0.007	0.631 ^b	0	0.390	-1.667	1.470	-0.003	0.936	4.13	0.20
ET-no q54, 68, 69	6	1193.72	0.05	0.008	0.784	0	0.641	-0.608	2.206	0.030	1.055	2.99	15.5 ^c

^aGlobal model fit was borderline in 3 dimensions.

^bPSI <0.70 in 6 dimensions.

^cMore than 5% of *t* statistics were significant in the final ET model.

Table 4 Summary of Item Importance, Psychometric Criteria, and Rasch Criteria from Final Models to Select 1 Item per Dimension

Dimension	Item Number and Topic	Patients—Rank Importance Based on Mean Item Importance (1 = Highest Ranked)		Clinicians—Rank Importance Based on Mean Item Importance (1 = Highest Ranked)		Floor and Ceiling Effects	% Missing Data	Correlation of Item to Dimension	Item Location	Range of Item Thresholds	Infit MSQ	Outfit MSQ	Differential Item Functioning ^a
		Rank	Importance	Rank	Importance								
Physical and role functioning	PF1. Trouble with strenuous activities	2	4 ^b	4 ^b	4 ^b		0	0.718	-0.794	4.358	0.879	0.876	
	PF2. Trouble taking a long walk	1	4 ^b	4 ^b	4 ^b		0.74	0.779	-0.262	3.793	0.749	0.738	X
	PF3. Trouble taking a short walk	6	3 ^c	3 ^c	3 ^c	Ceiling	0.25	0.604	2.079	2.657	0.763	0.497	
	PF4. Need to stay in a bed or chair	3	2	2	2		0.49	0.635	1.612	4.193	0.929	0.814	
	PF5. Need help eating, dressing, washing, using toilet	7	1	1	1	Ceiling	0	0.375	3.380	1.006	0.867	0.275	
Emotional functioning	During the past week...												
	PF6. Limited in doing work or other daily activities	4	3 ^c	3 ^c	3 ^c		0.25	0.741	0.363	3.622	0.874	0.840	
	FA10. Did you need to rest?	5	5	5	5		0.25	0.704	-0.567	5.601	0.854	0.850	
	During the past week...												
	EF21. Tense	2	4	4	4		0.25	0.749	1.051	5.874	0.763	0.745	
Social functioning	EF22. Worry	3 ^b	2	2	2		0.25	0.781	0.081	6.705	0.687	0.688	
	EF23. Irritable	3 ^b	3	3	3		0.25	0.784	1.247	5.226	0.789	0.741	
	EF24. Depressed	1	1	1	1		0.49	0.756	1.462	5.243	0.825	0.824	X
	During the past week...												
Fatigue	SF26. Physical condition or medical treatment interfered with family life	2	1	1	1		0.49	0.861	1.547	7.078	0.499	0.478	
	SF. 27. Physical condition or medical treatment interfered with social activities	1	2	2	2		0.49	0.861	0.872	7.433	0.499	0.466	
	During the past week...												
Fatigue	FA12. Felt weak	3	1	1	1		0	0.759	0.974	4.746	0.524	0.478	
	FA18. Tired	2	2	2	2		0.49	0.733	-0.635	7.015	0.563	0.542	
	ET56. Dizzy	1	3	3	3	ceiling	0	0.497	2.363	4.453	0.929	0.843	
During the past week...													

(continued)

Table 4 (continued)

Dimension	Item Number and Topic	Patients—Rank Importance Based on Mean Item Importance (1 = Highest Ranked)		Clinicians—Rank Importance Based on Mean Item Importance (1 = Highest Ranked)		Floor Ceiling and Effects	% Missing Data	Correlation of Item to Dimension	Item Location	Range of Item Thresholds	Infit MSQ	Outfit MSQ	Differential Item Functioning ^a
		1	2	1	2								
Pain	PA9. Had pain	1	2	2	2		0	0.792	0.688	7.571	0.543	0.523	
	PA19. Pain interfered with daily activities	2	1	1	1		0.98	0.792	1.55	5.859	0.489	0.495	
Body image	BI39: Physically less attractive	3	1	1	1		0.49	0.819	0.372	4.482	0.874	0.892	
	BI40: Less feminine	2	4	4	4		0.49	0.859	0.771	4.013	0.691	0.662	X
Sexual functioning and enjoyment	BI41: Problems looking at self naked	1 ^b	2	2	2	ceiling	0	0.850	1.273	3.996	0.774	0.773	
	BI42: Dissatisfied with your body	1 ^b	3	3	3		0.25	0.842	0.536	5.218	0.817	0.840	
Sexual functioning and enjoyment	During the past 4 weeks: SX.44 Have you been interested in sex?	1	1 ^b	1 ^b	1 ^b		3.19	0.783	1.012	5.016	0.713	0.731	
	SX.45 Have you been sexually active (with or without intercourse)?	2	2	2	2		3.43	0.747	1.367	5.873	0.783	0.812	
Systemic therapy side effects	SE.46. Has sex been enjoyable for you?	3	1 ^b	1 ^b	1 ^b		16.67	0.803	0.087	3.911	0.628	0.632	
	SYS 31. Dry mouth	4	4	4	4		0.507	0.509	0.172	1.714	0.895	0.901	X
Systemic therapy side effects	SYS 32. Food and drink tasted different than usual	2	5	5	5	Ceiling	0.598	0.59	1.454	N/A	0.722	0.565	
	SYS 33. Eyes painful, irritated or watery?	5	6	6	6		0.529	0.521	0.413	2.342	0.848	0.839	X
Systemic therapy side effects	SYS 34. Lost any hair?	1	2	2	2	Ceiling	0.458	0.457	0.499	N/A	0.954	0.906	
	SYS 36. Felt ill or unwell	6	1	1	1		0.612	0.621	0.059	1.518	0.799	0.788	
Systemic therapy side effects	SYS 38. Had headaches	3	3	3	3		0.321	0.336	-0.009	1.603	1.108	1.126	X

(continued)

Table 4 (continued)

Dimension	Item Number and Topic	Patients—		Clinicians—		Floor Ceiling and Effects	% Missing Data	Correlation of Item to Dimension	Item Location	Range of Item Thresholds	Infit MSQ	Outfit MSQ	Differential Item Functioning ^a
		Rank Importance Based on Mean Item Importance (1 = Highest Ranked)	Rank Importance Based on Mean Item Importance (1 = Highest Ranked)										
Arm and breast symptoms	ARM47. Pain in your arm or shoulder	3	5	0.25	0.692	-0.342	2.803	0.845	0.871				
	ARM48. Swollen arm or hand	7	2	0.25	0.548	0.466	1.602	1.074	1.216				
	ARM49. Problems raising your arm or moving it sideways	5	1	0.25	0.718	0.443	2.348	0.757	0.677				
	BR50. Pain in the area of your affected breast	6	6	0.25	0.743	0.218	3.139	0.730	0.756				
	BR51. Area of your affected breast swollen	1	4	0.49	0.614	0.952	1.676	0.903	0.857				
	BR52. Area of your affected breast been oversensitive	2	3 ^b	0.25	0.68	0.431	3.078	0.874	0.875				
	BR53. Skin problems on or in the area of your affected breast	4	3 ^b	0	0.549	0.539	1.598	1.069	1.181				
Endocrine therapy side effects	During the past week. . .	1	4	0.74	0.475	1.128	N/A	1.814	3.123				X
	YS37. Hot flushes	3	3 ^b	0.25	0.826	0.774	4.533	0.537	0.537				
	ET63. Problems with your joints	5	3 ^b	0	0.790	-0.212	2.500	0.669	0.683				
	ET64. Stiffness in your joints	2	1	0	0.867	-0.219	2.481	0.450	0.432				
	ET65. Pain in your joints	6	2	0	0.729	0.350	2.226	0.864	0.902				
	ET66. Aches or pains in your bones	4	3 ^b	1.23	0.680	1.234	4.798	0.959	1.024				
	ET67. Aches or pains in your muscles												

^aX is item with differential item functioning.^b tied rank importance

negligible in both items (likelihood ratio $\chi^2 P = 0.007$, McFadden's pseudo $R^2 = 0.006-0.020$). Removal of PF7 (*Hobbies*) improved the global model fit ($\chi^2 P = 0.29$) without degrading reliability significantly (new PSI = 0.776). PF2 (*Long walk*) had negligible DIF by age, with a significant likelihood ratio $\chi^2 P < 0.001$ comparing proportional odds models with and without the age parameter. For significantly different response categories, McFadden's pseudo R^2 was small (0.006–0.020).

PF2 (*Long walk*) was chosen to represent this dimension as it was rated most important by patients and tied fourth by clinicians, had the highest item-to-dimension correlation (0.779), had a moderately wide threshold range (3.855), and is clinically relevant based on clinician input.

Emotional functioning. The initial 5-item model from Rasch analysis had good global model fit ($\chi^2 P = 0.01$) and reliability (PSI = 0.774). All items had small mean standardized residuals (–0.006) and good infit and outfit. Two items had negligible DIF by age. EF24 (*Depressed*) had nonuniform DIF (likelihood ratio $\chi^2 P = 0.006$, McFadden's pseudo $R^2 = 0.004-0.013$), and ET55 (*Mood swings*) had uniform DIF (likelihood ratio $\chi^2 P < 0.001$, McFadden's pseudo $R^2 = 0.004-0.018$). Each item was iteratively removed and the Rasch model refitted. We chose the model without ET55 (*Mood swings*) because removing it had no effect on global model fit ($P = 0.01$) and a slight reduction in reliability (new PSI = 0.736). In contrast, removing EF24 (*Depressed*) further reduced model reliability (PSI = 0.719). All items had a high item-to-dimension correlation (>0.74), without floor or ceiling effects.

We chose EF22 (*Worry*) to represent emotional functioning because it was rated tied third most important by patients, second most important by clinicians, had the highest threshold range (6.705), and is clinically relevant.

Social functioning. The initial 2-item model had borderline global model fit ($\chi^2 P = 0.01$) and reliability (PSI = 0.603). Mean standardized item fit residuals were negligible (0.00). Both items SF26 (*Condition or treatment interfered with family life*) and SF27 (*Condition or treatment interfered with social activities*) had low infit (both 0.499) and outfit (0.478, 0.466, respectively) contributing to low global model fit. However, both items had a high item-to-dimension correlation (>0.86), without floor or ceiling effects. We chose SF27 (*Interfering with social activities*) because it was rated most important by

patients, second most important by clinicians, had the higher item threshold range (7.433), and is clinically relevant.

Fatigue. The initial 3-item dimension had good global model fit ($\chi^2 P = 0.02$) and suboptimal reliability (PSI = 0.526). Mean standardized item fit residuals were low (–0.015). Two items, FA12 (*Felt weak*) and FA18 (*Tired*), had low infit (0.524 and 0.563, respectively). Item ET56 (*Dizzy*) had ceiling effects (70.4%) and a low item-to-dimension correlation (0.497); therefore, this item was not chosen. We chose FA18 (*Tired*) to represent the Fatigue dimension because it was rated second most important by patients and clinicians alike, had the highest item threshold range (7.015), and is clinically relevant.

Pain. The initial 2-item dimension had suboptimal global model fit (PSI = 0.551) and reliability ($\chi^2 P = 0.01$). Mean standardized item fit residuals were low (–0.010), although infit and outfit of both items PA9 (*Had pain*) and PA19 (*Pain interfered with daily activities*) were suboptimal, with infit 0.688 and 1.55, respectively; outfit 0.523 and 0.495, respectively. Both items had a high item-to-dimension correlation (0.792), and neither item had floor or ceiling effects. We chose PA19 (*Had pain*) because it was rated most important by patients and second most important by clinicians, had the largest item threshold range (7.571), and is clinically relevant.

Body image. This initial 4-item dimension had good global model fit ($\chi^2 P = 0.01$) and reliability (PSI = 0.767). All items had small mean standardized item fit residuals (–0.007) and high item-to-dimensional correlation (>0.81). All except for 1 item had good infit and outfit: BI40 (*Less feminine*) had borderline infit (0.691) and outfit (0.662). BI40 (*Less feminine*) also had uniform DIF by age, although the magnitude was negligible (likelihood ratio $\chi^2 P < 0.001$, McFadden's pseudo $R^2 = 0.0049-0.020$). BI41 (*Problems looking at yourself naked*) had ceiling effects (50.9%). We chose BI42 (*Dissatisfied with your body*) because it was rated tied most important by patients and third most important by clinicians, had the largest item threshold range (5.218), and is clinically relevant.

Sexual functioning and enjoyment. This initial 3-item dimension had borderline global model fit ($\chi^2 P = 0.01$) and suboptimal reliability (PSI = 0.543). Mean

standardized item fit residuals were low (0.001), although SE 46 (*Has sex been enjoyable?*) had borderline low infit (0.628) and outfit (0.632). All items correlated highly with the dimension (>0.74), and none had floor or ceiling effects. We chose SX 44 (*Interest in sex*) because it was ranked most important by patients and clinicians and had the second largest range of item thresholds (5.016). This item is the most inclusive and clinically relevant in this dimension, since interest in sex does not assume sexual intercourse.

Systemic therapy side effects. This 6-item dimension had good global model fit ($\chi^2 P = 0.01$) but poor reliability (PSI = 0.355). Reliability improved slightly, although it remained poor (new PSI = 0.392) after recoding response options of 4 items: SYS32 (*Food and drink tasted different*), SYS34 (*Lost hair*), SYS36 (*Felt ill or unwell*), and SYS38 (*Headaches*). Two dichotomous items had ceiling effects: SYS32 (*Food and drink*; 90.4%) and SYS34 (*Lost hair*; 80.3%). Mean item standardized fit residuals were low (-0.012), and item fit was good, except that SYS32 (*Food and drink*) had low outfit (0.565). Three items had DIF of low magnitude: SYS31 (*Dry mouth*) had nonuniform DIF (likelihood ratio $\chi^2 P < 0.001$, McFadden's pseudo $R^2 = 0.008-0.018$), and both SYS33 (*Eyes painful, irritated, or watery*) and SYS38 (*Headaches*) had uniform DIF (likelihood ratio $\chi^2 P < 0.001$ and $P < 0.001$, respectively; McFadden's pseudo $R^2 = 0.003-0.025$ and $0.011-0.029$, respectively). All items had overall low item-to-dimension correlations (0.336–0.621). For this dimension, we chose SYS34 (*Lost hair*), which was recoded as dichotomous based on overlapping response options, with a point threshold at 0.499. SYS34 (*Lost hair*) was ranked highest among patients, ranked second among clinicians, and is clinically relevant.

Arm and breast symptoms. This 7-item dimension had good global model fit ($\chi^2 P = 0.06$), and its reliability was borderline (PSI = 0.631). Item fit had low mean standardized fit residuals (-0.003) and good infit and outfit, except that ARM49 (*Problems raising your arm or moving it sideways*) had borderline outfit (0.677). Four items had ceiling effects: ARM48 (*Swollen arm or hand*; 69.3%), ARM49 (*Problems raising arm*; 62.4%), BR51 (*Area of your affected breast was swollen*; 74.1%), and BR53 (*Skin problems on affected breast*; 68.4%). After recoding ordered, overlapping response options of ARM48 (*Swollen arm or hand*), BR50 (*Pain of your affected breast*), BR51 (*Affected breast oversensitive*),

BR53 (*Skin problems on the effected breast*), discrimination worsened (new PSI = 0.594); therefore, we kept the original response options. We chose BR52 (*Affected breast oversensitive*) because it had the second highest ranked importance by patients, was ranked third by clinicians, had an acceptable item-to-dimension correlation (0.68), had the second largest threshold range (3.078), and is clinically relevant.

Endocrine therapy symptoms. This 9-item dimension had initial good global fit ($\chi^2 P = 0.05$) and reliability (PSI = 0.760). Mean item standardized fit residuals was low (0.30). Three items were iteratively removed because they had high infit (>1.3): ET54 (*Sweated excessively*), ET68 (*Gained weight*), and ET69 (*Weight gain a problem*). This resulted in improved reliability (PSI = 0.784). SYS37 (*Hot flushes*), assigned to endocrine therapy from the confirmatory factor analysis,²⁵ was retained because it was rated most important by patients despite its high infit (1.81) and outfit (3.12) and uniform DIF (likelihood ratio $\chi^2 P = 0-0.419$, McFadden's pseudo $R^2 = 0.001-0.023$). Three items with low infit and outfit (<0.70) were retained ET63 (*Problems with your joints*), ET64 (*Stiffness in your joints*), ET65 (*Pain in your joints*) since iterative removal lowered model reliability (PSI < 0.7). No items had floor or ceiling effects. Although a PCA of item residuals revealed no factors, 15.5% of t statistics were significant. We chose ET63 (*Problems with your joints*) because it had the second highest range of item thresholds (4.533), with a high item-to-dimension correlation (0.826), was ranked third most important by patients and clinicians, and is clinically relevant.

Table 4 presents a summary of the item selection criteria for each dimension. Table 5 presents the item selected to represent each dimension.

Discussion

The contribution of this study is the identification of 1 representative item per dimension for the future BUI. The BUI will be a novel BrC-specific preference-based instrument derived from the EORTC QLQ-C30 and BR45. The representative items were chosen based on a priori psychometric and Rasch criteria and investigator judgment. We chose the following 10 items: 1) Physical functioning (*Trouble taking a long walk*), 2) Emotional functioning (*Worry*), 3) Social functioning (*Interfering with social activities*), 4) Pain (*Having pain*), 5) Fatigue (*Tired*), 6) Body image (*Dissatisfied with your body*), 7) Systemic therapy side effects (*Hair loss*), 8) Sexual functioning (*Interest in sex*), 9) Breast symptoms

Table 5 Breast Utility Instrument Items Selected from the EORTC QLQ-C30 and BR45 Dimensions Using a Priori Criteria

QLQ-C30	BR45
Physical and role functioning PF2. Trouble taking a long walk	Body image During the past week: BI42. Dissatisfied with your body.
Emotional functioning During the past week: EF22. Did you worry?	Systemic therapy side effects During the past week: SYS34. Lost any hair?
Social functioning During the past week: SF27. Physical condition or medical treatment interfered with your social activities.	Sexual functioning and enjoyment During the past week: SX44. Interested in sex?
Pain During the past week: PA9. Did you have pain?	Breast and arm symptoms During the past week: BR52. Affected breast been oversensitive?
Fatigue During the past week: FA18. Have you felt tired?	Endocrine therapy symptoms During the past week: ET63. Problems with your joints.

(*Oversensitive breast*), and 10) Endocrine therapy symptoms (*Problems with your joints*).

Comparing our selected items with the general cancer preference-based instruments EORTC-8D¹⁵ and QLU-C10D,¹⁶ 4 items were the same and 6 were different. Four items that we selected were included in the QLU-C10D: PF2 (*Long walk*), SF27 (*Interfered with your social activities*), PA9 (*Pain*), and FA18 (*Tired*). The EORTC-8D selected PA19 (*Pain interfered with daily activities*). Both QLU-C10D and EORTC-8D selected EF24 (*Depressed*), while we chose EF22 (*Worry*) for the emotional functioning dimension. These generic cancer preference-based instruments covered 8 and 10 dimensions, respectively, so their items spanned a wider range of cancer-specific symptoms, including nausea, and constipation or diarrhea. The QLU-C10D included an item for trouble sleeping. Unlike previous studies, this study selected representative items from BrC-specific dimensions from the BrC-specific QLQ-BR45.

Our approaches differed based on our derivation samples and study design. The EORTC-8D was derived from patients with multiple myeloma,¹⁵ and the QLU-C10D was derived from diverse cancer sites from European countries.¹⁶ Both instruments included item responsiveness from longitudinal clinical trial data as item selection criteria, which was inherently not possible with our cross-sectional sample.

While some researchers prefer generic preference-based instruments over CSPBI, there is a compelling reason for a CSPBI in BrC. CSPBIs can arguably create classification systems that are difficult to compare because of excluded side effects, focusing effects, and naming effects, and comorbidities may influence

valuation of the health state condition.¹³ The BUI, a future CSPBI derived from the EORTC QLQ-C30 and BR45, however, can potentially better reflect preference-sensitive outcomes affecting body image, hair loss, loss of interest in sex, breast sensitivity, or problems with the joints, content not directly captured in generic preference-based instruments.

Our procedure for selecting items for the BUI has several strengths. We applied clinimetrics, an approach from clinical medicine,⁸¹ by drawing on the judgment of patients and clinicians using their item importance ratings as criteria for item selection. We also applied psychometrics, mathematical techniques that originated in psychology and education,⁸² with decades of applications in health instrument development. After fitting the Rasch model, we considered the range of item thresholds across the latent construct, thus improving precision.⁸³ These approaches were also adopted by developers of other CSPBIs: FACT-8D for general cancer,⁸⁴ CORE-6D for mental health,⁸⁵ CP-6D for cerebral palsy,⁸⁶ and CARIES-QC for dental caries.³³ Our approach could also be applied to the development of other generic preference-based instruments.

There are several limitations to this study. All participants attended medical oncology clinics from 1 tertiary cancer center, limiting generalizability. Future validation samples should accrue patients undergoing a wider range of treatments, across all 5 health states, and a variety of locations. There are limitations to the item importance approach, which relies on the judgment of a subset of patients and clinicians with respect to the importance associated with each item, either from their individual perspective (patients) or according to patients

(clinicians). The sequence of applying the item importance approach could have influenced the final selection of items. Patients and clinicians were asked to complete item importance ratings prior to our confirmatory factor analysis,²⁵ and the ranking of mean item importance ratings were considered in our final selection of each item per dimension.

Conclusions



The proposed 10 items for the BUI are parsimonious dimensions of a novel BrC-specific preference-based instrument. Reducing QLQ-C30 and BR45 items to 10 items will greatly reduce respondent burden during clinical practice, clinical trials, and future studies that administer multiple concurrent questionnaires. The 10 items of the BUI will be amenable to valuation in future studies towards developing the BUI.

Future development of the BUI will include assessing measurement properties (reliability, criterion and construct validity, and responsiveness) prior to eliciting direct utility weights (e.g., time trade off) from patients and community members and modeling utilities of health states as a function of health state dimensions (attributes).⁸⁷ The future BUI potentially will better allow women's preferences to be reflected in clinical decision making and cost utility analyses.

Acknowledgments

We are grateful for the contributions of the following individuals who made this research possible: Dr. Sofia Torres laid the groundwork for developing a breast cancer preference-based instrument. Dr. Andrea Eisen, Dr. Kataryna Jerzak, Dr. Rosanna Pezo, Dr. Ellen Warner, and Dr. Sonal Ghandi kindly allowed us to recruit their patients during their busy clinics. Dr. Donilo Giffoni, Elizabeth Matheson, Kim Nguyen, Dr. Neda Stepanovic, Lisa Verity, and the M6 nursing staff took time out of their busy schedules to help create opportunities for patient interactions. Doyoung (Kelly) Kim assisted with patient recruitment and data entry. Arcturus Phoon assisted with chart abstraction. We also thank the Biomatrix team, Nim Li, Cordelia He, and Dr. Martin Yaffe, for ensuring that our data was well captured and secure.

ORCID iDs

Teresa C. O. Tsui  <https://orcid.org/0000-0002-9806-9393>
 Nicholas Mitsakakis  <https://orcid.org/0000-0001-9041-9152>

Supplemental Material

Supplementary material for this article is available on the *MDM Policy & Practice* website at <https://journals.sagepub.com/home/mpp>.

References

1. Aaronson NK, Ahmedzai S, Bergman B, et al. The European organization 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;85:365–76.
2. Bjelic-Radisic V, Cardoso F, Cameron D, et al.; EORTC Quality of Life Group and Breast Cancer Group. An international update of the EORTC questionnaire for assessing quality of life in breast cancer patients: EORTC QLQ-BR45. *Ann Oncol.* 2020;31:283–8. DOI: 10.1016/j.annonc.2019.10.027
3. Wolowacz SE, Briggs A, Belozeroff V, et al. Estimating health-state utility for economic models in clinical studies: an ISPOR good research practices task force report. *Value Health.* 2016;19:704–19. DOI: 10.1016/j.jval.2016.06.001
4. Husereau D, Drummond M, Petrou S, et al.; CHEERS Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *Value Health.* 2013;16:e1–5. DOI: 10.1016/j.jval.2013.02.010
5. Fisher ES, Wennberg JE. Health care quality, geographic variations, and the challenge of supply-sensitive care. *Perspect Biol Med.* 2003;46:69–79.
6. Keirns CC, Goold SD. Patient-centered care and preference-sensitive decision making. *JAMA.* 2009;302:1805–6. DOI: 10.1001/jama.2009.1550
7. Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. *JAMA.* 2008;300:436–8. DOI: 10.1001/jama.300.4.436
8. Dolan P. Modeling valuations for EuroQol health states. *Med Care.* 1997;35:1095–108.
9. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population survey. Discussion Paper 138. University of York; 1995.
10. Teckle P, Peacock S, McTaggart-Cowan H, et al. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. *Health Qual Life Outcomes.* 2011;9:106. DOI: 10.1186/1477-7525-9-106
11. Rowen D, Brazier J, Ara R, Azzabi Zouraq I. The role of condition-specific preference-based measures in health technology assessment. *Pharmacoeconomics.* 2017;35(suppl 1):33–41. DOI: 10.1007/s40273-017-0546-9
12. Goodwin E, Green C. A systematic review of the literature on the development of condition-specific preference-based measures of health. *Appl Health Econ Health Policy.* 2016;14:161–83. DOI: 10.1007/s40258-015-0219-9
13. Brazier J, Rowen D, Mavranzouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess.* 2012;16:1–114. DOI: 10.3310/hta16320
14. Davis AM, Perruccio AV, Canizares M, et al. The development of a short measure of physical function for hip OA HOOS-Physical Function Shortform (HOOS-PS): an

- OARSI/OMERACT initiative. *Osteoarth Cartil.* 2008;16:551–9. DOI: 10.1016/j.joca.2007.12.016
15. Rowen D, Brazier J, Young T, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health.* 2011;14:721–31.
 16. King MT, Costa DS, Aaronson NK, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30 [published corrected appears in *Qual Life Res.* 2016;25[10]:2683]. *Qual Life Res.* 2016;25:625–36.
 17. Collins ED, Moore CP, Clay KF, et al. Can women with early-stage breast cancer make an informed decision for mastectomy? *J Clin Oncol.* 2009;27:519–25. DOI: 10.1200/JCO.2008.16.6215
 18. Potter S, Brigid A, Whiting PF, et al. Reporting clinical outcomes of breast reconstruction: a systematic review. *J Natl Cancer Inst.* 2011;103:31–46. DOI: 10.1093/jnci/djq438
 19. Tran NV, Chang DW, Gupta A, Kroll SS, Robb GL. Comparison of immediate and delayed free TRAM flap breast reconstruction in patients receiving postmastectomy radiation therapy. *Plast Reconstr Surg.* 2001;108:78–82. DOI: 10.1097/00006534-200107000-00013
 20. Kunneman M, Engelhardt EG, Ten Hove FL, et al. Deciding about (neo-)adjuvant rectal and breast cancer treatment: missed opportunities for shared decision making. *Acta Oncol.* 2016;55:134–9. DOI: 10.3109/0284186X.2015.1068447
 21. Preminger BA, Pusic AL, McCarthy CM, Verma N, Worku A, Cordeiro PG. How should quality-of-life data be incorporated into a cost analysis of breast reconstruction? A consideration of implant versus free TRAM flap procedures. *Plast Reconstr Surg.* 2008;121:1075–82. DOI: 10.1097/01.prs.0000304246.66477.cd
 22. Goldhirsch A, Gelber RD, Simes RJ, Glasziou P, Coates AS. Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis. *J Clin Oncol.* 1989;7:36–44. DOI: 10.1200/JCO.1989.7.1.36
 23. Hillner BE, Weeks JC, Desch CE, Smith TJ. Pamidronate in prevention of bone complications in metastatic breast cancer: a cost-effectiveness analysis. *J Clin Oncol.* 2000;18:72–9. DOI: 10.1200/JCO.2000.18.1.72
 24. Tsui TCO, Torres S, Bielecki J, et al. Developing the Breast Utility Instrument to Assess Health-related Quality of Life and Preferences in Patients with Breast Cancer. June 2022. <https://tspace.library.utoronto.ca/handle/1807/123349> (accessed 29 November 2022).
 25. Tsui TCO, Trudeau M, Mitsakakis N, et al. Developing the breast utility instrument, a preference-based instrument to measure health-related quality of life in women with breast cancer: confirmatory factor analysis of the EORTC QLQ-C30 and BR45 to establish dimensions. *PLoS One.* 2022;17:e0262635. DOI: 10.1371/journal.pone.0262635
 26. Juniper EF, Guyatt GH, Streiner DL, King DR. Clinical impact versus factor analysis for quality of life questionnaire construction. *J Clin Epidemiol.* 1997;50:233–8.
 27. Sprangers MA, Groenvold M, Arraras JI, et al. The european organization for research and treatment of cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *J Clin Oncol.* 1996;14:2756–68. DOI: 10.1200/JCO.1996.14.10.2756
 28. Lidgren M, Wilking N, Jonsson B, et al. Health related quality of life in different states of breast cancer. *Qual Life Res.* 2007;16:1073–81. DOI: 10.1007/s11136-007-9202-8
 29. Lidgren M, Wilking N, Jönsson B, Rehnberg C. Resource use and costs associated with different states of breast cancer. *Int J Technol Assess Health Care.* 2007;23:223–31. DOI: 10.1017/S0266462307070328
 30. Krahn M, Ritvo P, Irvine J, et al. Construction of the patient-oriented prostate utility scale (PORPUS): a multi-attribute health state classification system for prostate cancer. *J Clin Epidemiol.* 2000;53:920–30.
 31. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum.* 2007;57:1358–62. DOI: 10.1002/art.23108
 32. Bray N, Tudor Edwards R. Preference-based measurement of mobility-related quality of life: developing the Mob-QoL-7D health state classification system. *Disabil Rehabil.* 2022;44:2915–29. DOI: 10.1080/09638288.2020.1844319
 33. Rogers HJ, Gilchrist F, Marshman Z, Rodd HD, Rowen D. Selection and validation of a classification system for a child-centred preference-based measure of oral health-related quality of life specific to dental caries. *J Patient Rep Outcomes.* 2020;4:105. DOI: 10.1186/s41687-020-00268-9
 34. Matza LS, Phillips G, Dewitt B, et al. A scoring algorithm for deriving utility values from the Neuro-QoL for patients with multiple sclerosis. *Med Decis Making.* 2020;40:897–911. DOI: 10.1177/0272989X20951782
 35. Keetharuth AD, Rowen D, Björner JB, Brazier J. Estimating a preference-based index for mental health from the recovering quality of life measure: valuation of recovering quality of life utility index. *Value Health.* 2021;24:281–90. DOI: 10.1016/j.jval.2020.10.012
 36. Kerr C, Breheny K, Lloyd A, et al. Developing a utility index for the Aberrant Behavior Checklist (ABC-C) for fragile X syndrome. *Qual Life Res.* 2015;24:305–14. DOI: 10.1007/s11136-014-0759-8
 37. Mulhern B, Rowen D, Jacoby A, et al. The development of a QALY measure for epilepsy: NEWQOL-6D. *Epilepsy Behav.* 2012;24:36–43.
 38. Mulhern B, Smith SC, Rowen D, et al. Improving the measurement of QALYs in dementia: developing patient- and carer-reported health state classification systems using Rasch analysis. *Value Health.* 2012;15:323–33.
 39. Rowen D, Mulhern B, Banerjee S, et al. Comparison of general population, patient, and carer utility values for dementia health states. *Med Decis Making.* 2015;35:68–80. DOI: 10.1177/0272989X14557178
 40. Nguyen KH, Mulhern B, Kularatna S, Byrnes J, Moyle W, Comans T. Developing a dementia-specific health state classification system for a new preference-based instrument

- AD-5D. *Health Qual Life Outcomes*. 2017;15(1):21. DOI: 10.1186/s12955-017-0585-0
41. Comans TA, Nguyen KH, Mulhern B, et al. Developing a dementia-specific preference-based quality of life measure (AD-5D) in Australia: a valuation study protocol. *BMJ Open*. 2018;8:e018996. DOI: 10.1136/bmjopen-2017-018996
 42. Mavranetzouli I, Brazier JE, Rowen D, Barkham M. Estimating a preference-based index from the Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM): valuation of CORE-6D. *Med Decis Making*. 2013;33:381–95. DOI: 10.1177/0272989X12464431
 43. Mavranetzouli I, Brazier JE, Young TA, Barkham M. Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). *Qual Life Res*. 2011;20:321–33. DOI: 10.1007/s11136-010-9768-4
 44. Cuervo J, Castejón N, Khalaf KM, Waweru C, Globe D, Patrick DL. Development of the incontinence utility index: estimating population-based utilities associated with urinary problems from the incontinence quality of life questionnaire and neurogenic module. *Health Qual Life Outcomes*. 2014;12:147. DOI: 10.1186/s12955-014-0147-7
 45. Castejón N, Khalaf K, Ni Q, Cuervo J, Patrick DL. Psychometric properties of the incontinence utility index among patients with idiopathic overactive bladder: data from two multicenter, double-blind, randomized, Phase 3, placebo-controlled clinical trials. *Health Qual Life Outcomes*. 2015;13:116. DOI: 10.1186/s12955-015-0306-5
 46. Oluboyede Y, Hulme C, Hill A. Development and refinement of the WAIte: a new obesity-specific quality of life measure for adolescents. *Qual Life Res*. 2017;26:2025–39.
 47. Mukuria C, Rowen D, Brazier JE, Young TA, Nafees B. Deriving a preference-based measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value Health*. 2015;18:846–55. DOI: 10.1016/j.jval.2015.07.004
 48. Dzingina M, Higginson IJ, McCrone P, Murtagh FEM. Development of a patient-reported palliative care-specific health classification system: the POS-E. *Patient*. 2017;10:353–65.
 49. Goodwin E, Green C. A quality-adjusted life-year measure for multiple sclerosis: developing a patient-reported health state classification system for a multiple sclerosis-specific preference-based measure. *Value Health*. 2015;18:1016–24. DOI: 10.1016/j.jval.2015.07.002
 50. 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;31:195–210. DOI: 10.1177/0272989X10364846
 51. Mulhern B, Labeit A, Rowen D, et al. Developing preference-based measures for diabetes: DHP-3D and DHP-5D. *Diabet Med*. 2017;34:1264–75.
 52. Kowalski JW, Rentz AM, Walt JG, et al. Rasch analysis in the development of a simplified version of the national eye institute visual-function questionnaire-25 for utility estimation. *Qual Life Res*. 2012;21:323–34.
 53. Rentz AM, Kowalski JW, Walt JG, et al. Development of a preference-based index from the national eye institute visual function questionnaire-25. *JAMA Ophthalmol*. 2014;132:310–8.
 54. Goh RLZ, Fenwick E, Skalicky SE. The visual function questionnaire: utility index: does it measure glaucoma-related preference-based Status? *J Glaucoma*. 2016;25:822–9. DOI: 10.1097/IJG.0000000000000441
 55. Misajon R, Hawthorne G, Richardson J, et al. Vision and quality of life: the development of a utility measure. *Invest Ophthalmol Vis Sci*. 2005;46:4007–15. DOI: 10.1167/iovs.04-1389
 56. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-specific preference-based measures: benefit or burden? *Value Health*. 2012;15:504–13. DOI: 10.1016/j.jval.2011.12.003
 57. Kuspinar A, Finch L, Pickard S, Mayo NE. Using existing data to identify candidate items for a health state classification system in multiple sclerosis. *Qual Life Res*. 2014;23:1445–57. DOI: 10.1007/s11136-013-0604-5
 58. Sundaram M, Smith MJ, Revicki DA, Elswick B, Miller LA. Rasch analysis informed the development of a classification system for a diabetes-specific preference-based measure of health. *J Clin Epidemiol*. 2009;62:845–56. DOI: 10.1016/j.jclinepi.2009.01.020
 59. Young T, 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;18:253–65. DOI: 10.1007/s11136-008-9428-0
 60. Yang Y, Brazier J, Tsuchiya A, Coyne K. Estimating a preference-based single index from the overactive bladder questionnaire. *Value Health*. 2009;12:159–66. DOI: 10.1111/j.1524-4733.2008.00413.x
 61. Chiou C-F, Suarez-Almazor ME, Sherbourne CD, et al. Development and validation of a preference weight multi-attribute health outcome measure for rheumatoid arthritis. *J Rheumatol*. 2006;33:2409–11.
 62. Dobrez D, Cella D, Pickard AS, Lai JS, Nickolov A. Estimation of patient preference-based utility weights from the functional assessment of cancer therapy—general. *Value Health*. 2007;10:266–72.
 63. Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications in test development. *Educ Meas Issues Pract*. 1993;12:38–47.
 64. Masters G. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47:149–74.
 65. Brown TA. *Confirmatory Factory Analysis for Applied Research*. New York: Guilford Press; 2006.
 66. Christensen KB, Kreiner S, Mesbah M. *Rasch Models in Health*. Hoboken (NJ): John Wiley & Sons; 2013.
 67. Choi SW, Gibbons LE, Crane PK. lordif: logistic ordinal regression differential item functioning using IRT. R package.

- 0.3-3 ed. 2016. Available from: <https://CRAN.R-project.org/package=lordif> (accessed 2 May 2021)
68. 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. DOI: 10.1348/014466506x96931
 69. O'Brien RM. Identification of simple measurement models with multiple latent variables and correlated errors. *Sociol Methodol.* 1994;24:137–70.
 70. Team RC. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing; 2021. Available from: <https://www.R-project.org/> (accessed 2 May 2021)
 71. Revelle W. *psych: Procedures for Personality and Psychological Research.* 2.1.3 ed. Evanston (IL): Northwestern University; 2020. Available from: <https://CRAN.R-project.org/package=psych> (accessed 1 June 2021)
 72. Mair P, Hatzinger R, Maier MJE. *eRm: extended Rasch modeling.* 2021. Available from: <https://cran.r-project.org/package=eRm> (accessed 6 April 2021)
 73. Rizopoulos D. ltm: an R package for latent variable modelling and item response theory analyses. *J Stat Softw.* 2006;17:1–25.
 74. Linacre JM. Sample size and item calibration stability. *Rasch Meas Trans.* 1994;7:328.
 75. Perlis N, Krahn MD, Boehme KE, et al. The bladder utility symptom scale: a novel patient reported outcome instrument for bladder cancer. *J Urol.* 2018;200:283–91. DOI: 10.1016/j.juro.2018.03.006
 76. Perlis N, Krahn M, Alibhai S, et al. Conceptualizing global health-related quality of life in bladder cancer. *Qual Life Res.* 2014;23:2153–67. DOI: 10.1007/s11136-014-0685-9
 77. Seung SJ, Traore AN, Pourmirza B, Fathers KE, Coombes M, Jerzak KJ. A population-based analysis of breast cancer incidence and survival by subtype in Ontario women. *Curr Oncol.* 2020;27:e191–8. DOI: 10.3747/co.27.5769
 78. Canadian Cancer Statistics Advisory Committee. Canadian cancer statistics. 2018. Available from: cancer.ca/Canadian-Cancer-Statistics-2018-EN.pdf. Accessed August 3, 2021.
 79. Statistics Canada. Census profile, 2016 census. Catalogue No. 98-316-X2016001. 2017. Available from: <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E>. Accessed August 4, 2021.
 80. Mair P. *Modern Psychometrics With R.* Cham (UK): Springer; 2018.
 81. Feinstein A. The theory and evaluation of sensibility. In: Feinstein A, ed. *Clinometrics.* New Haven (CT): Yale University Press; 1987. p 141–65.
 82. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development and Use.* Oxford (UK): Oxford University Press; 2008.
 83. Clark LA, Watson D. Constructing validity: basic issues in objective scale development. *Psychol Assess.* 1995;7:309–19.
 84. King MT, Norman R, Mercieca-Bebber R, et al.; Multi-attribute utility in cancer consortium. The Functional Assessment of Cancer Therapy Eight Dimension (FACT-8D), a multi-attribute utility instrument derived from the cancer-specific FACT-General (FACT-G) quality of life questionnaire: development and Australian value set. *Value Health.* 2021;24:862–73. DOI: 10.1016/j.jval.2021.01.007
 85. Mavranzouli I, Brazier JE, Rowen D, Barkham M. Estimating a preference-based index from the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM): valuation of CORE-6D. *Med Decis Making.* 2013;33:381–95. DOI: 10.1177/0272989X12464431
 86. Bahrapour M, Downes M, Boyd RN, Scuffham PA, Byrnes J. Using Rasch and factor analysis to develop a proxy-reported health state classification (descriptive) system for cerebral palsy. *Disabil Rehabil.* 2021;43:2647–55. DOI: 10.1080/09638288.2019.1709565
 87. Brazier JE, Rowen D, Mavranzouli I, et al. Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol Assess.* 2012;16:1–114.