Health and Quality of Life Outcomes

REVIEW

Open Access

A scoping review to create a framework for the steps in developing condition-specific preference-based instruments de novo or from an existing non-preference-based instrument: use of item response theory or Rasch analysis

Teresa C. O. Tsui[1,2,3*], Sofia C. Torres[4,5], Joanna M. Bielecki[1], Nicholas Mitsakakis[1,6], Maureen E. Trudeau[7], Karen E. Bremner[1], Aileen M. Davis[4†] and Murray D. Krahn[1†^]

Abstract

Background There is no widely accepted framework to guide the development of condition-specific preference-based instruments (CSPBIs) that includes both de novo and from existing non-preference-based instruments. The purpose of this study was to address this gap by reviewing the published literature on CSPBIs, with particular attention to the application of item response theory (IRT) and Rasch analysis in their development.

Methods A scoping review of the literature covering the concepts of all phases of CSPBI development and evaluation was performed from MEDLINE, Embase, PsychInfo, CINAHL, and the Cochrane Library, from inception to December 30, 2022.

Results The titles and abstracts of 1,967 unique references were reviewed. After retrieving and reviewing 154 full-text articles, data were extracted from 109 articles, representing 41 CSPBIs covering 21 diseases or conditions. The development of CSPBIs was conceptualized as a 15-step framework, covering four phases: 1) develop initial questionnaire items (when no suitable non-preference-based instrument exists), 2) establish the dimensional structure, 3) reduce items per dimension, 4) value and model health state utilities. Thirty-nine instruments used a type of Rasch model and two instruments used IRT models in phase 3.

Conclusion We present an expanded framework that outlines the development of CSPBIs, both from existing non-preference-based instruments and de novo when no suitable non-preference-based instrument exists, using IRT

[†]Aileen M. Davis and Murray D. Krahn are joint senior authors.

We dedicate this manuscript to Dr. Murray Dale Krahn who passed away on July 1, 2022, during its preparation. We are grateful for Dr. Krahn's insight about the need to outline the steps of condition-specific preference-based instrument development. This manuscript is one product of his legacy.

*Correspondence:
Teresa C. O. Tsui
teresa.tsui@utoronto.ca
Full list of author information is available at the end of the article

and Rasch analysis. For items that fit the Rasch model, developers selected one item per dimension and explored item response level reduction. This framework will guide researchers who are developing or assessing CSPBIs.

Keywords Scoping review, Health-related quality of life, Condition-specific preference-based instrument, Framework, Rasch analysis, Item response theory

Plain language summary

Preference-based instruments are questionnaires that generate “utility”, a global measure of quality of life anchored at 0 for “dead” to 1 for “full health”. Condition-specific preference-based instruments assess utility on dimensions that are relevant to individuals living with specific health problems. We reviewed the published literature on condition-specific preference-based instruments to develop a framework of steps to guide and standardize their development. Our focus was the use of Rasch analysis and item response theory, statistical methods to select the core questions. Using these methods, the resulting instruments are more precise as fewer questions can represent the key dimensions.

We found 109 articles which described the development of 41 condition-specific preference-based instruments for 21 different diseases or conditions. The development of condition-specific preference-based instruments was conceptualized as a 15-step framework, covering four phases: 1) develop initial questionnaire items, 2) establish the dimensions of the items, 3) reduce the number of items per dimension, and 4) value and model health state utilities. These phases extend from the development of the non-preference-based instrument and multi-attribute classification system to the preference-based instrument. This framework will improve the development of instruments to measure utilities for specific conditions, which will enable the use of high-quality utilities in clinical decisions and cost-effectiveness analyses.

Introduction

Condition-specific preference-based instruments (CSPBI) measure health-related quality of life (HRQoL) relevant to patients with a specific condition or disease. In contrast, generic preference-based instruments such as the EQ-5D family of questionnaires [1] are suitable for general use [1–3]. Preference-based instruments contain a classification system with items representing attributes and levels within items which, with a value set, produce a utility score anchored at zero (dead) and one (perfect health). Values are derived from patients or members of the general public who provided utilities for health states using direct methods, including time trade off (TTO) [3, 4] or discrete choice experiments (DCE) [3, 4]. Utility is used to calculate the quality-adjusted life year (QALY), a key outcome in economic evaluations and clinical decision-making.

Guidelines recognize that CSPBI discriminate between known groups better than generic instruments and are more responsive to changes in disease-specific dimensions [5–9].

Several systematic reviews included aspects of CSPBI development, including one that found 51 different CSPBIs [5, 10–13]. Brazier et al. [5] described six stages of preference-based instrument development starting with an existing condition-specific non-preference-based instrument, such as the Functional Assessment of Cancer Therapy – General measure (FACT-G) scale [14], or European Organization for Research and Treatment of Cancer (QLQ-C30) [15] in oncology. The stages are: I) establish dimensionality, II) select items for each dimension, III) test the number of levels, IV) validate the health state classification system, V) valuation survey, and VI) model the valuation data. When there is no established condition-specific non-preference-based instrument, the steps in the development of a CSPBI begin with creating a classification system of domains de novo [13, 16].

Factor analysis (confirmatory or exploratory) is used to establish dimensions. Item response theory (IRT) or Rasch analysis can be used to eliminate items and select one or two items to represent each dimension [5]. Item response theory (IRT) is a measurement approach that explains the probabilistic relationship between items and a latent construct (e.g., HRQoL) [17]. The Rasch model is the simplest IRT model [18]. When items fit the Rasch model, the instrument has favourable properties: unidimensionality, interval-level scoring, additivity, and sample-free measurement [19]. Instruments developed with Rasch or IRT methods have high precision and efficiency by selecting the fewest items to cover the latent construct [19, 20]. Health states are then sampled and modelled using a decomposed or composite approach [5]. While these stages provide a starting point for the development of novel CSPBI, the methods described by Brazier et al. begin with an existing condition-specific non-preference-based instrument. Within these stages, there are insufficient details for novice CSPBI developers to follow. Additionally, when there is no suitable condition-specific non-preference based instrument, developing a novel CSPBI de novo is the best option. These initial steps of creating a non-preference-based instrument de novo have been described by Guyatt et al. [16], yet these steps were absent from the Brazier et al. stages [5].

The aim for this scoping review is to address these gaps by operationalizing Brazier et al.'s stages based on available literature, and adding the initial steps to develop a preference-based instrument de novo when there is no existing HRQoL instrument. Our focus was the use of Rasch and IRT methods to establish dimensions, reduce items per dimension, and reduce item levels because resulting instruments have favourable properties. These steps underpin the creation of a multi-attribute health state classification system to develop a novel preference-based instrument. Our objectives were to:

1. Identify the steps in constructing CSPBIs, both de novo and from an existing non-preference-based instrument.
2. Describe the application of Rasch or IRT methods within these steps.
3. Develop an expanded framework to guide future development of CSPBIs.

Methods

Information sources

We followed the Joanna Briggs Institute (JBI) published guidance document [21, 22], and the Preferred Reporting Items for Systematic Review and Meta-Analysis Scoping Review (PRISMA-ScR) reporting guidelines (Supplementary Information, S1) [23]. Our scoping review protocol was not published.

Searches were performed in Ovid MEDLINE, Ovid Embase, Ovid PsychInfo, EBSCO CINAHL, and the Cochrane Library from inception to December 2022 (Supplementary Information, S2). An experienced health sciences librarian (JB) and TT developed a search strategy (Supplementary Information, S2) using Medical Subject Headings (MeSH) and keywords about:

1. Measurement of condition-specific HRQoL
2. Eliciting health state utility values to develop a preference-based instrument
3. Methods to develop instruments measuring HRQoL
4. IRT including Rasch analysis

The search strategy was reviewed by a second librarian, following the Peer Review of Electronic Search Strategies (PRESS) standard [24].

Selection of articles

Search results were imported into Thomson Reuters EndNote X9.3.3 to remove duplicates.

A primary (TT) and secondary (ST) reviewer independently screened titles and abstracts, followed by full text articles using Covidence [25]. We excluded abstracts, commentaries, editorials, letters, and non-English articles. Articles were excluded if they predicted utilities from only demographics or other non-disease factors, or validated non-English instruments, since these do not describe the development of the instruments.

We included articles that described either the development of a CSPBI using IRT or Rasch analysis, or the elicitation of utility weights for the instrument. Articles about instruments had the following measurement purposes: 1) to discriminate between known disease states, or 2) to measure responsiveness after treatment and over time.

We also hand-searched the reference list of Goodwin's systematic review [13] for the names of instruments. Additional searches were performed using individual instrument names on Pubmed from inception to February 2024 (Supplementary Information, S3). We chose the review by Goodwin and Green because it included all steps of CSPBI development, and was the most recent and most comprehensive of the review papers that we found.

Inter-reviewer reliability was assessed using a kappa statistic, with cut-off scores: 0.40–0.59 for *fair* agreement, 0.60–0.74 for *good* agreement, and 0.75 and higher for *excellent* agreement [26]. Discrepancies in interpreting eligibility criteria were discussed, and the criteria were revised for clarity if inter-reviewer reliability was below good [26].

Data extraction

The steps of instrument development were extracted from full text articles. The data extraction form (Supplementary Information S4) was pilot tested on 10 articles that covered all instrument development phases and was iteratively revised until it captured all essential steps. One reviewer (TT) extracted the data from all articles and a second reviewer (ST) reviewed the data against all articles. Discrepancies were resolved by discussion.

Constructing the framework

We started with Brazier et al.'s six stages outlining how to derive CSPBIs from existing psychometric instruments [5]. Next, we reviewed existing frameworks for the development of classification systems of domains for non-preference-based instruments [16, 27], and for use of factor analysis [28, 29] and Rasch analysis [19]. Finally, we reviewed articles describing the development of CSPBIs to identify the key steps.

Results

Study selection

Figure 1 shows the PRISMA diagram. After removal of duplicates, the titles and abstracts of 1,967 references were reviewed, and 71 additional references were identified from hand-searching Goodwin's systematic

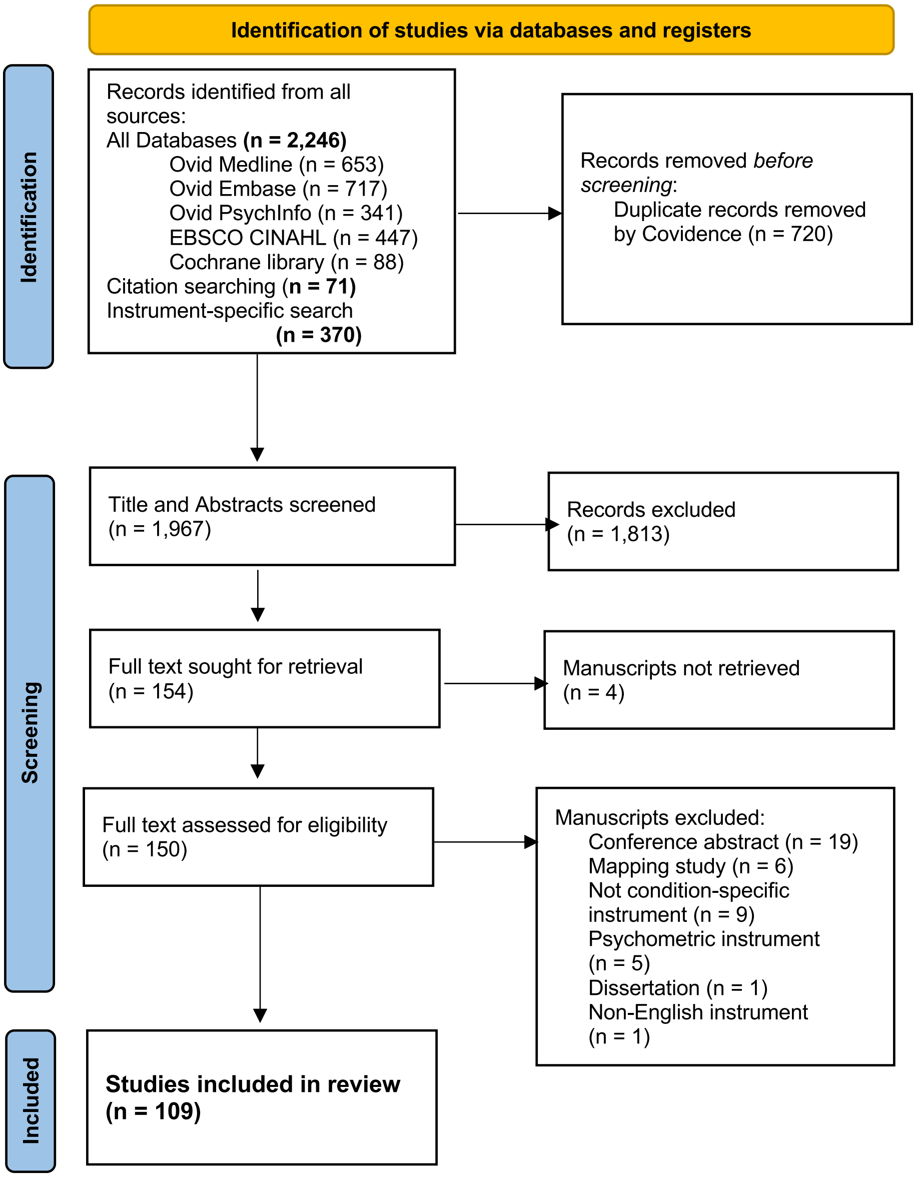

Identification of studies via databases and registers
Identification
Records identified from all sources:
All Databases (n = 2,246)
Ovid Medline (n = 653)
Ovid Embase (n = 717)
Ovid PsychInfo (n = 341)
EBSCO CINAHL (n = 447)
Cochrane library (n = 88)
Citation searching (n = 71)
Instrument-specific search (n = 370)
Records removed before screening:
Duplicate records removed by Covidence (n = 720)
Screening
Title and Abstracts screened (n = 1,967)
Records excluded (n = 1,813)
Full text sought for retrieval (n = 154)
Manuscripts not retrieved (n = 4)
Full text assessed for eligibility (n = 150)
Manuscripts excluded:
Conference abstract (n = 19)
Mapping study (n = 6)
Not condition-specific instrument (n = 9)
Psychometric instrument (n = 5)
Dissertation (n = 1)
Non-English instrument (n = 1)
Included
Studies included in review (n = 109)

Fig. 1 PRISMA Diagram [30]

review [13]. One hundred and fifty-four full-text articles were retrieved and reviewed. Data were extracted from 109 articles representing 41 unique instruments, and 21 diseases/conditions. Inter-rater agreement was fair (kappa = 0.57) during the title and abstract screening, and good (kappa = 0.71) during the full text selection.

Study characteristics

Table 1 outlines the 41 CSPBIs, which covered 21 diseases or conditions. Instruments had 4 to 10 dimensions (median = 6). Only the classification systems were developed for three instruments, without preference elicitation. Direct preference elicitation methods were used for 38 instruments: TTO (22), TTO and rating scale (RS) (1), DCE (1), DCE with TTO (DCE$_{\text{-TTO}}$) (5), DCE and DCE$_{\text{-TTO}}$ (1), DCE$_{\text{-TTO}}$ and best worst scaling (2), standard gamble (SG) (1), SG and RS (1), SG and VAS (1), and VAS alone (1). Sources of utilities (number of studies) were the general public only (22), patients only (7), general public and patients (7), general public and patient and care-giver dyads (1), and general public and carers (1) (Table 1).

Phases and steps to developing CSPBIs

Figure 2 shows the framework of the four phases and 15 steps of CSPBI development.

Table 1 Published condition-specific preference-based instruments

Condition	Instrument [references]	Dimensions	Number of health states	Method of eliciting utilities	Groups for utility elicitation	Scaling anchors	Country
Aberrant behaviour	Aberrant Behavior Checklist -Utility Index (ABC-UI) [31, 32]	7	3^(7) = 2,187	TTO (LT-TTO worse than dead)	UK general public	Best(1), worst (0)	UK
Asthma	Asthma Quality of Life Questionnaire- 5 Dimensions (AQL-5D) [33–39]	5	5^(5) = 3,125	TTO	UK general public	Best (1), worst (0)	UK
Bladder problems	Incontinence Utility Index (IUI) [40, 41]	5	3^(5) = 243	TTO	UK general public	Perfect health (1), dead (0)	UK Spain
Bladder problems	Overactive Bladder Questionnaire- 5 Dimensions (OAB-5D) [9, 36, 42–44]	5	5^(5) = 3,125	TTO	UK general public	Full health (1), dead (0)	UK USA
Cancer	European Organisation for Research and Treatment in Cancer Core Quality of Life Questionnaire – 8 Dimensions (EORTC-8D) [7, 45–48]	8	81,920	TTO	UK general public Sri Lankan general public	Best (1), worst (0)	UK Sri Lanka
Cancer	European Organisation for Research and Treatment in Cancer Core Quality of Life Questionnaire – 10 Dimensions QLU-C10D [49–60]	10	4^(10) = 1,048,576	DCE-$_{TTO}$ (Australia, France, Austria, Italy, and Poland) DCE (Canada, Netherlands, Germany, UK, U.S.)	General public	Best (1), worst (0)	Australia Netherlands Canada France Germany UK USA Austria, Italy, and Poland
Cancer	European Organisation for Research and Treatment in Cancer Core Quality of Life Questionnaire – Preference-Based Measure (QLQ-PBM) [61]	8	105	TTO	Dutch general public	Best (1),worst (-1) -1 to 0 worse than dead	Netherlands
Cancer	Functional Assessment of Cancer Therapy – 8 Dimensions (FACT-8D) [62, 63]	8	1744 participants; 16 choice pairs; 256 observations per choice set	DCE	Patients with common cancers	Best (1), worst (0)	Australia

Table 1 (continued)

Condition	Instrument [references]	Dimensions	Number of health states	Method of eliciting utilities	Groups for utility elicitation	Scaling anchors	Country
Cancer (breast)	Breast Utility Instrument (BUI) [64, 65]	10	N/A	N/A	N/A No. utilities were elicited yet, so patients with breast cancer is an error.	N/A	Canada
Cerebral palsy	Cerebral Palsy-6 Dimensions (CP-6D) [66–68]	6	5^(6) = 15,625	DCE-$_{\text{TTO}}$	Australian general public	Best (1), worst (0)	Australia
Cystic fibrosis	Cystic Fibrosis Questionnaire-Revised-8 Dimensions (CFQ-R-8 D) [69]	8	3^(8) = 6,561	TTO	UK general public	Best (1), worse than dead (< 0)	UK
Dementia	Dementia Quality of Life-Utility (patient self-report and carer proxy-report) DEMQOL-U DEMQOL-Proxy-U [8, 70–74]	5 4	4^(5) = 1,024 4^(4) = 256	TTO	UK general public and patients UK general public and carers	Best (1), worst (0)	UK
Dementia	Alzheimer's Disease-5 Dimensions (AD-5D) [75–77]	5	4^(5) = 1,024	DCE$_{\text{-TTO}}$ BWS	Australian general public and patients with dementia & carer dyads	Best (1), worst (0)	Australia
Diabetes	Diabetes Utility Index (DUI) [78, 79]	5	4^(5) = 1024	VAS SG	People with type 1 or 2 diabetes	Best (1), worst (0)	USA
Diabetes	Diabetes Health Profile-3 Dimensions and 5-Dimensions (DHP-3D and DHP-5D) [80]	3 5	33 were valued	TTO	UK general public	Best (1), worst (0)	UK
Diabetes	Health and Self-Management in Diabetes Questionnaire (HASMID) (HASMID-8 HASMID-10) [81, 82]	8 10	262,144 profiles 120 choice sets across 10 survey versions	DCE-$_{\text{TTO}}$	UK general public and Individuals with diabetes	Unanchored and WTP estimates	UK
Duchenne Muscular Dystrophy	Duchenne Muscular Dystrophy Quality of Life-8 Dimensions (DMD-QoL-8D) [83]	8	4^(8) = 65,536	DCE-$_{\text{TTO}}$	UK general public	Best (1), dead (0)	UK

Table 1 (continued)

Condition	Instrument [references]	Dimensions	Number of health states	Method of eliciting utilities	Groups for utility elicitation	Scaling anchors	Country
Epilepsy	Epilepsy-specific Quality of Life – 6 Dimensions (NEWQOL-6D) [84–86]	6	4^(6) = 4,096	TTO	UK general public People with epilepsy	Best (1), worst (0)	UK
Heart disease	MacNew Heart Disease Health-Related Quality of Life Instrument -7 Dimensions (MacNew) [87]	7	4^(7) = 16,384	N/A	N/A	N/A	Australia
Human immunodeficiency virus (HIV)	Preference-based HIV index (PB-HIV)	7	3^(7) = 2,187	EQ-VAS	People with HIV	Best (100), and worst (0)	Canada
Mental health	Clinical Outcomes in Routine Evaluation-Outcome Measure – 6 Dimensions (CORE-6D) [88–90]	6	33 plausible health states (Rasch vignette approach) 3^(6) = 729 CORE-6D health states	TTO	UK general public	Full health (1), dead (0)	UK
Mental health	Recovering Quality of Life Utility Index (ReQoL-UI) [91–94]	7	66 health states valued Each participant valued 8 health states	TTO Lead-time TTO (states worse than dead)	Nationally representative sample from Scotland, England, Wales	Best (1), worst (0)	UK
Mobility	Mobility Quality of Life-7 Dimensions (MobQoL-7D) [95, 96]	7	4^(7) = 16,384	VAS	UK general public Representative sample of people with impaired mobility	Best (100), worst (0)	UK
Multiple sclerosis	Multiple Sclerosis Impact Scale – 8 Dimensions, and -8 Dimensions patient versions (MSIS-8D MSIS-8D-P) [27, 97–100]	8	4^(8) = 65,536	TTO	UK general public People with MS	Best (1), worst (0)	UK
Multiple sclerosis	Multiple Sclerosis Impact Scale – Preference-Based Measure (MSIS-PBM) [61]	6	100	TTO	Dutch general public	Best (1), worst (-1) -1 to 0 worse than dead	Netherlands
Multiple sclerosis	Health-related quality of life in people with neurological conditions (Neuro-QoL-Utility System (NQU)) [101, 102]	6	59 health states 47 single-attribute states 6 corner states 3 marker states: worst, best, dead	SG	General public sample, and patients with MS	Best (1), worst (0)	UK

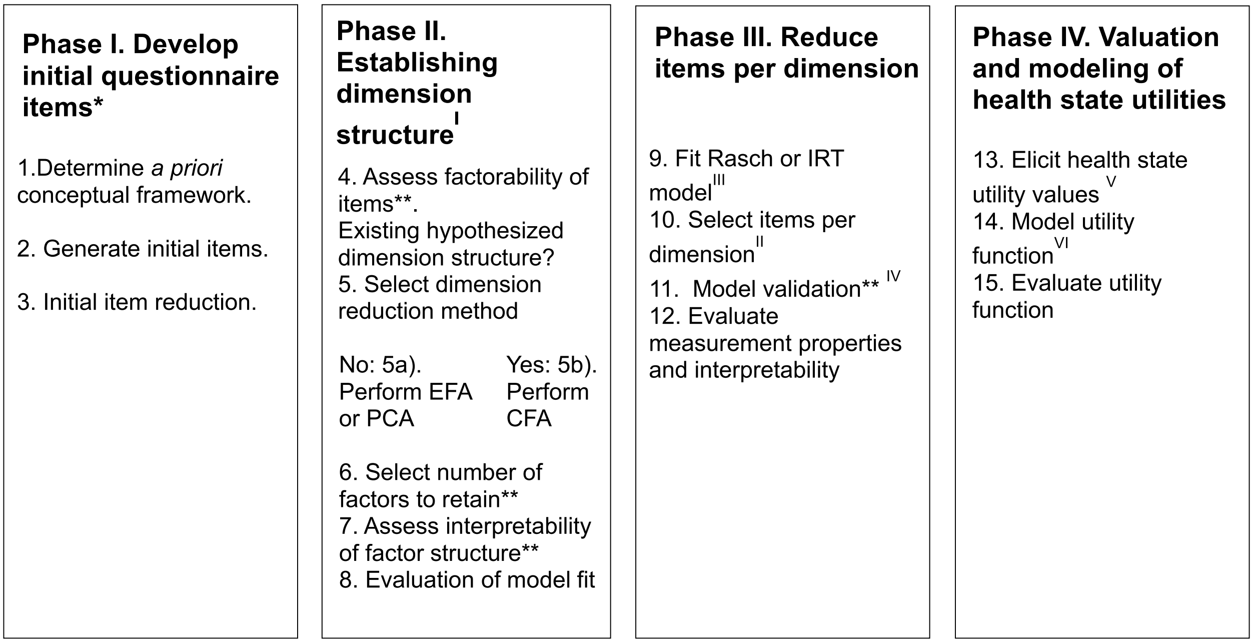
Table 1 (continued)

Condition	Instrument [references]	Dimensions	Number of health states	Method of eliciting utilities	Groups for utility elicitation	Scaling anchors	Country
Multiple sclerosis	Preference-Based Multiple Sclerosis Index (P-PBMSI) [103–106]	5	3^(5) = 243	SG RS	People with MS	Best (1), worst (0)	Canada
Myelofibrosis	Myelofibrosis-8 Dimensions (MF-8D) [107]	8	2560	TTO	UK general public	Best (1), worst (0)	UK
Obesity	Weight-specific Adolescent Instrument for Economic-Evaluation (WAITe) [108–110]	5	7^(5) = 16,807	TTO	UK general public	Best (1), worst (0)	UK
Obesity	Preference-Based Index of Weight-Related Quality of Life (PBI-WRQL) [111]	7	3^(7) = 2,187	EQ-VAS	Patients	Best (100), worst (0)	Canada
Oral health	Caries Impacts and Experiences Questionnaire for Children Utility version (CARIES-QC-U) [112, 113]	4	3^(5) = 243	DCE$_{\text{-TTO}}$ BWS	Adolescents and adults from the general public	Best (1), worst (0)	UK
Oral health	Early Childhood Oral Health Impact Scale-4 Dimensions (ECOHIS-4D) [114, 115]	4	3^(4) = 81	DCE$_{\text{-TTO}}$	Australian general public	Best (1), worst (0)	Australia
Palliative care	Palliative Care Outcome Scale Descriptive System (POS-E) [116]	3	3^(7) = 2187	N/A	N/A	N/A	UK
Rheumatoid arthritis	Multiattribute Health Outcome Measure for Rheumatoid Arthritis (MHOM RA) [117]	6	4^(6) = 40,096	TTO (general health preference)	Patients with rheumatoid arthritis	Best (1), worst (0)	UK
Rheumatoid arthritis	Health Assessment Questionnaire – Preference-Based Measure (HAQ-PBM) [61]	5	56	TTO	Dutch general public	Best (1), worst (0)	Netherlands

Table 1 (continued)

Condition	Instrument [references]	Dimensions	Number of health states	Method of eliciting utilities	Groups for utility elicitation	Scaling anchors	Country
Vision	Vision Quality of Life Index (Vis-QoL) [118–124]	6	5^(6) =15,625	TTO and RS (intermediate item responses)	Patients with visual impairment (Canada) Patients with diabetic retinopathy, diabetic macular edema, and keratoconus (Australia) Patients with visual impairment (India)	Best (1), dead (0), States worse than dead -0.25	Canada Australia India
Vision	Visual Function Questionnaire – Utility Index (VFQ-UI) [125–130]	6	5^(6) = 15,625	TTO	General public from Australia, Canada, UK, USA	Best (1), worst (0)	Australia, Canada, UK, USA

Abbreviations: *TTO* Time trade-off, *DCE-TTO* Discrete-choice experiment, time-trade-off, *LT-TTO* Lead time-time trade-off, *BWS* Best worst scaling, *RS* Rating scale, *VAS* Visual analogue scale, *UK* United Kingdom, *USA* United States of America



Phase I. Develop initial questionnaire items*

1.Determine *a priori* conceptual framework.

2. Generate initial items.

3. Initial item reduction.

Phase II. Establishing dimension structure[I]

4. Assess factorability of items**.
Existing hypothesized dimension structure?
5. Select dimension reduction method

No: 5a). Perform EFA or PCA

Yes: 5b). Perform CFA

6. Select number of factors to retain**
7. Assess interpretability of factor structure**
8. Evaluation of model fit

Phase III. Reduce items per dimension

9. Fit Rasch or IRT model[III]
10. Select items per dimension[II]
11. Model validation**[IV]
12. Evaluate measurement properties and interpretability

Phase IV. Valuation and modeling of health state utilities

13. Elicit health state utility values[V]
14. Model utility function[VI]
15. Evaluate utility function

* Phase I (steps 1 to 3) is required when there is no existing non-preference-based instrument.
**Phase II (steps 4, 6, and 7) are desirable when performing EFA or PCA. Phase III. (step 11) model validation is desirable.
Brazier's stages: I: Establish dimensions. II: Eliminate and select items per dimension and III: Explore item level reduction. IV: Validation – repeat stages I to III on other data sets. V: Health state selection and valuation study. VI: Modelling to produce preference weights for all states.

Fig. 2 Framework

Phase I (Steps 1–3): Conceptualize measurement construct and develop initial items

These three initial steps were conducted for the 7 instruments developed de novo (Table 2). These steps are only required when developing a CSPBI de novo and therefore are absent from Brazier's stages, which start with an existing non-preference-based instrument. The data to gather for phase I are the relevant literature of frameworks and existing items, and results from patient interviews or focus groups.

Step 1. Determine a priori conceptual framework A conceptual framework defines the construct to be measured. The purpose of starting with a conceptual framework with defined core dimensions is to ensure that measurement of the construct is comprehensive and has established boundaries [132, 133]. Three instruments were developed with condition-specific conceptual frameworks (DUI, WAITe, Vis-QoL) [78, 108, 118]. Frequently, developers reviewed the literature [70, 75, 103, 134] and conducted focus groups to create an a priori framework [70, 91, 118]. Other developers analyzed literature and interviews using content analysis [70], grounded theory [118], or framework analysis [92, 93] (Table 2).

Step 2. Generate initial items The purpose of generating an initial comprehensive pool of items is to cover the entire construct to be measured [20]. Items that represented the conceptual framework of the descriptive system were generated using literature reviews (WAITe) [108], chart reviews, or other existing HRQoL instruments [31, 78, 103, 117, 125, 134, 135]. Patient and clinician experts were consulted in interviews (WAITe) [108], and focus groups (MHOM RA, VisQoL) [42, 117, 118] (Table 2), which consider patient perspectives [136].

Step 3. Initial item reduction The purpose of initial item reduction is to ensure alignment of the items with an a priori framework of HRQoL [103] (Table 2), and to remove redundant items [20]. Developers field-tested the VisQoL in people with and without vision impairment [118, 126]. Developers reduced items after consultation with patients, carers, and/or clinicians (MHOM RA, WAITe, Vis-QoL) [108, 117, 118], performing framework analysis (WAITe) [108]. Development of the PBI-WRQL [111] and PB-HIV [131] removed correlated dimensions (r > / = 0.3) and mapped initial items to an established framework to establish the instrument dimensions (Table 2).

Table 2 Phase I (Steps 1–3) Conceptualize measurement construct and develop initial items

	HIV PB-HIV [131]	**Multiple sclerosis P-PBMSI [103–106]**	**Obesity WAITe [108, 109]**	**Obesity PBI-WRQL [111]**	**Rheumatoid arthritis MHOM RA [117]**	**Vision Vis-QoL [118–124]**	**Diabetes Utility Instrument DUI [78]**
Step 1: Determine a priori conceptual framework							
Conceptual framework			X			X	X
Literature review, existing instruments, chart records							X
Gather expert opinions						P	R, HCP
Analysis method						Grounded theory	
Step 2: Generate initial items							
Literature review, existing instruments		X	X		X		
Gather expert opinions			P		HCP	P	
Analysis method			Identify themes based on framework				
Step 3: Initial item reduction							
Conceptual framework	X	X		X			
Gather expert opinions	P, HCP		P	P		P	
Analysis method	Removed correlated dimensions; alignment to framework using mapping		Alignment to framework	Removed correlated dimensions; alignment to frame-work using mapping			

P Patients, *G* General public, *C* Carers, *HCP* Health care providers, *R* Researchers

Phase II (Steps 4–8): Establishing dimension structure

Factor and principal component analyses (PCA) are data aggregation techniques that explain the pattern of correlations between items and latent constructs, such as HRQoL dimensions [28] (Table 3). Phase II overlaps with Brazier's stage I (Fig. 2). The intent of establishing the dimensional structure is to assess structural independence, which means there is a low correlation between dimensions [137]. The data to collect for phase II are responses to the questionnaire.

Step 4. Assess factorability of items The factorability of items indicates whether it is feasible to proceed with factor analysis [28]. Coefficients of 0.3 to 0.8 in a correlation matrix [31, 108, 112] or > 0.70 in Cronbach's alpha [81, 118] are criteria for factorability. If performing PCA or exploratory factor analysis (EFA), developers also assessed the Bartlett test of sphericity, and Kaiser-Meyer Olkin measure of sampling adequacy [31, 66, 87, 108, 116, 141, 142] (Table 3).

Step 5. Select dimension extraction method Next, developers chose a dimension extraction method. The first consideration is whether or not a hypothesized dimension structure exists (Fig. 2). Given an a priori dimensional structure (e.g., a psychometric instrument), authors performed CFA to test the hypothesis and fit covariances or correlations between items and factors (ReQol-UI, FACT-8D) [62, 94] (Table 3). Without an a priori hypothesis of the dimensional structure, most developers performed EFA (CFQ-8D, CP-6D, DEMQOL-U, and DHP-3D and 5D) [66, 69, 71, 80] or PCA (ABC-UI, AQL-5D, IUI, OAB-5D, CORE-6D, POS-E, MHOM RA) [31, 33, 40, 42, 88, 116, 117]. Some authors performed EFA and then CFA (AD-5D, WAITe, QLU-C10D, and DUI) [75, 78, 108, 141], or vice-versa (DMD-QoL) [134]. PCA was used to

Table 3 Phase II (Steps 4 to 8) Establish the dimension structure

	ABC-UI [31, 32]	AQL-5D [33–39]	IUI [40, 41]	OAB-5D [9, 36, 42–44]	EORTC-8D [7, 45–48]	QLU-C10D [49–58]	HAQ-PBM, QLQ-PBM, MSIS-PBM [61]	FACT-8D [62, 63]	CP-6D [66]	DEMQOL-U & -Proxy-U [8, 71–74]	AD-5D [75–77]	CARIES-QC-U [112]	DUI [78, 79]	DHP-3D & 5D [80]	HAS-MID-8 & 10 [81, 82]	DMD-QoL-8D [134, 138, 139]
Step 4. Assess factorability of items																
Item-total correlations$^{E, C, Pa}$	X						X					X				
Bartlett test of sphericity$^{E, Pb}$						X			X							
Kaiser-Meyer Olkin measure proportion of variance$^{E, Pc}$						X			X							
Cronbach's alpha^{d}			X												X	
Step 5. Select dimension extraction method																
Principal components analysis (PCA)	X	X	X	X												
Exploratory factor analysis (EFA)									X	X				X		
Confirmatory factor analysis (CFA)								X								
EFA and CFA						X					X		X			X
Step 6. If there is no hypothesized dimensional structure, select the factors to retain																
Cattell scree test$^{\mathbf{E, P}}$				X		X										
Amount of variance explained$^{\mathbf{E, P}}$ – eigen values		X	X	X	X				X					X		
Parallel analysis$^{\mathbf{E, P}}$						X										

Table 3 (continued)

Step 7. Assess interpretability of dimension structure																	
Oblimin rotation						X											
Promax rotation									X								
Varimax rotation					X					X							
Varimax and promax rotation	X																
Step 8. Evaluate model fit																	
RMSEA^{+}						X					X					X	
SRMR^{+}																	
CFI^{++}						X					X					X	
TLI^{++}						X					X						
Factor loadings				X	X	X	X				X		X	X			
Residual correlations						X											
Cross loading	X	X			X	X							X				
	NEWQoL-6D [84–86]	CORE-6D [88–90]	ReQoL-UI [91–94]	MobQoL-7D [95, 140]	MSIS-8D & -P [27, 97–100]	Neuro-QoL derived NQU [101, 102]	P-PBMSI [103–106]	MF-8D [107]	WAITe [108, 109]	POS-E [116]	MHOMRA [117]	Vis-QoL [118–124]	VFQ-UI [125–130]	Mac-New-7D [87]	ECO-HIS-4D [114]	BUI [64]	CFQ-8D [69]
Step 4. Assess factorability of items																	
Item-total correlations$^{E, C, Pa}$									X							X	X
Bartlett test of sphericity$^{E, Pb}$				X						X				X	X		
Kaiser-Meyer Olkin measure proportion of variance$^{E, Pc}$				X						X				X	X		
Cronbach's alpha^{d}				X							X	X			X	X	

Table 3 (continued)

Step 5. Select dimension reduction method														
Principal components analysis (PCA)		X						X	X		X			
Exploratory factor analysis (EFA)	X	X		X	X	X				X		X		X
Confirma-tory factor analysis (CFA)			X										X	
EFA and CFA							X							
Cattell scree test$^{E, P}$		X		X	X	X					X	X		X
Step 6. If there is no hypothesized dimensional structure, select the factors to retain														
Amount of variance explained$^{E, P}$ – eigen values			X	X	X						X	X		X
Parallel analysis$^{E, P}$							X	X			X	X		
Step 7. Assess interpretability of dimension structure														
Oblimin rotation														
Promax rotation		X		X					X					
Varimax rotation					X			X				X		
Varimax and promax rotation											X			
8. Evaluate model fit														
RMSEA^{+}			X							X			X	
SRMR^{+}													X	
CFI^{++}			X							X			X	
TLI^{++}													X	

Table 3 (continued)

Factor loadings	X	X	X	X	X	**X**	X	X
Cross loading								
Residual correlations							X	

E: exploratory factor analysis, P: principal components analysis, C: confirmatory factor analysis

Factor loading: > 0.3 or > 0.4. + root mean square error of approximation (RMSEA) and standardized root mean squared residual (SRMR) < 0.08 acceptable, < 0.05 good, + + comparative fit index (CFI) and Tucker-Lewis index (TLI) > 0.9 acceptable, > 0.95 good

[a] Inter-item correlations range: 0.3–0.8

[b] Bartlett's test of sphericity: Chi-sq, *p*-value < 0.01, < 0.05, confirm presence of correlations among items, to indicate whether EFA would be plausible

[c] KMO: > 0.5, or > 0.8, close to 1.0

[d] Cronbach's alpha: > 0.7 for group use, > 0.8 for individual use

reduce a set of variables to a smaller set of components [29] (Fig. 2).

Step 6. Select the number of factors to retain If there was no hypothesized dimensional structure, developers had to decide on the number of factors to retain to best represent the underlying structure of the dataset [28]. In PCA and EFA, developers considered the amount of variance that was explained by the eigen values [33, 45, 80, 87] or visualized in a scree plot [69, 87, 88, 107, 141], or they performed parallel analysis to interpret the scree plots more objectively [87, 108, 116, 141] (Table 3).

Step 7. Assess interpretability of dimension structure Within PCA and EFA, developers assessed the factor-loading matrix for interpretability, or meaning [28]. If necessary, developers improved interpretability using: i) promax and oblimin rotation to produce correlated factors (CORE-6D, MHOM RA, QLU-C10D) [49, 88, 117], and ii) varimax rotation to produce uncorrelated factors (EORTC-8D, DEMQOL-U, MSIS-8D, and POS-E) [45, 70, 97, 116]. Developers of the ABC-UI and MacNew-7D used both types of rotation [31, 87]. These methods of rotation help to achieve a structurally simpler matrix than the original factor loading matrix [28] (Table 3).

Step 8. Evaluation of model fit The purpose of evaluating model fit is to assess whether the model needs revision to fit the data. Developers evaluated global model fit using root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) (<0.08 acceptable, <0.05 good), and comparative fit index (CFI) and Tucker-Lewis index (TLI) (>0.9 acceptable, >0.95 good) [49, 64, 75, 94, 118, 134]. Developers evaluated factor loadings (>0.3 or >0.4) to ensure the item loaded sufficiently to the factor. In PCA and EFA, developers considered cross-loading differences (<0.15, or <0.2) to assign the item to the dimension with the higher loading (ABC-UI, AQL-5D, EORTC-8D, QLU-C10D, DUI) [31, 33, 45, 78, 141]. If model fit was inadequate using any data aggregation approach, developers re-inspected factor loadings and applied residual correlations to improve overall global fit (e.g.,QLU-C10D, BUI) [49, 64]. Developers of the DMD-QoL found poor initial fit using CFA, but fit was improved in a 3-dimensional hierarchical model using EFA [134] (Table 3).

Phase III (Steps 9–13): Reducing items per dimension

Together with Phase II, the purpose of reducing items per dimension in Phase III is to create a preference-based instrument that is amenable to valuation [3]. The data required to perform Phase III are responses to the questionnaire, which can be the same set of data used in Phase II.

Step 9. Fit Rasch or IRT model Rasch and IRT models have different purposes, originating from two diverging traditions. Rasch models belong to a model-based tradition since the model is selected first, and the tests are designed to determine if the data fit the model. Proponents of the Rasch model posit that the Rasch model represents the structure of item responses before they can be used for measurement [143]. In the alternative data-based traditions, different models within the IRT family are explored to find the best fitting model for the available data [144].

Thirty-nine of 41 instruments fit the data to a Rasch model. Six instruments used a Rasch rating scale model, nine used the Rasch partial credit model, and 24 used an unspecified polytymous model. Two instruments fit an IRT graded response model (GRM) (ReQoL-UI, NQU) [94, 101] (Table 4).

Aligned with Brazier's stage III (explore item level reduction) [5] (Fig. 2), CSPBI developers who conducted Rasch analysis first evaluated item response ordering to collapse disordered categories, or removed items with disordered response options, and re-ran the model. Sometimes developers asked experts to review the language of merged categories for clarity and comprehensiveness (face validity) [84, 88, 116] (Table 4).

Developers who used Rasch analysis then evaluated model fit, item fit, and person fit. Global model fit was assessed with an item-trait interaction χ^2 (non-significant, with Bonferroni correction) and/or person separation index, similar to Cronbach's alpha or reliability (>0.7, or >0.8) [27, 33, 61, 70, 84, 88, 103, 108, 116, 117, 125, 145]. Many developers reported item and then person fit statistics [31, 40, 78]. Mean item fit residuals and mean person fit residuals, measures of divergence between expected and observed responses for item or person responses, respectively, were evaluated. Residuals >2.5 or <-2.5 represent poor fit [27, 33, 61, 70, 84, 88, 103, 108, 116, 125]. Additional chi-square statistics were used to investigate observed vs expected responses for items with a severity level near the person's HRQoL level (infit) or for all items (outfit) [66, 78, 83, 95, 142], where a significant chi-square means an item misfits the model [19] (Table 4).

Next, some developers assessed local dependence and/or differential item functioning (DIF) to explain poor item fit. Local dependence occurs when the response to one item is linked to another item, evaluated by examining a residual correlation matrix [45, 49, 65, 70, 75, 108, 116]. Locally dependent items, such as trouble taking a short walk and trouble taking a long walk, were combined

Table 4 Phase III (Steps 9 to 11) Reducing items per dimension

	ABC-UI [31, 32]	AQL-5D [33–39]	IUI [40, 41]	OAB-5D [9, 36, 42–44]	EORTC-8D [7, 45–48]	QLU-C10D [49–58]	BUI [131]	HAQ-PBM, QLQ-PBM, MSIS-PB [61]	FACT-8D [62, 63]	CP-6D [66]	DEMQOL-U & -Proxy-U [8, 71–74]	AD-5D [75–77]	CAR-IES-QC-U [112]	DUI [78, 79]	DHP-3D & 5D [80]	HAS-MID-8 & 10 [81, 82]	DMD-QoL [134, 138, 139]
Step 9. Select Rasch or IRT model and fit model																	
Rasch rating scale				X										X			X
Rasch partial credit		X	X				X			X							
Unspecified polytymous Rasch	X				X	X		X	X		X	X	X		X	X	
Graded response IRT																	
Response level ordering[a]		X	X	X	X	X	X	X	X		X	X	X		X	X	
Meaningfulness of merged levels																	
Item parameters at logit 0	X		X	X	X	X	X	X	X					X			
Global model fit – item-trait interaction[b]		X		X	X	X	X	X		X	X		X				
Person separation index or person separation reliability[c]		X		X	X		X					X		X			
Infit, Outfit							X			X				X			X
Item fit residuals[d]		X		X		X	X	X	X		X	X	X				
Local dependence						X	X		X								
Differential item functioning	X		X	X	X	X	X	X	X	X		X	X		X	X	X

Table 4 (continued)

Person fit residuals[e]					X		X					X					
Targeting of scale to persons																	
Unidimensionality – principal components analysis of residuals, independent t-statistics			X								X			X			
Step 10. Select items per dimension																	
Item coverage across range of construct						X	X		X	X			X				X
Floor and ceiling effects				X	X	X	X	X			X		X			X	X
Missing data (%)	X			X	X		X	X					X				X
Correlation of item to dimension					X												X
Item importance and impact[f]	HCP					P	P, HCP		P								X
Step 11. Model validation																	
Expert validation										X			X	X			
Another data-set or split half		X		X													
Meaningful clinical rationale (face and content validity[f]					HCP	HCP, P		HCP, R		HCP, R		R	P, G	HCP, P	HCP		

Table 4 (continued)

Alignment with parent psychometric instrument[C, E, P]					X	X												
	NEWQoL-6D [84–86]	CORE-6D [88–90]	ReQoL-UI [91–94]	MobQoL-7D [95, 140]	MSIS-8D & -P [27, 97–100]	Neuro-QoL derived NQU [101, 102]	P-PBMSI [103–106]	MF-8D [107]	WAITe [108, 109]	POS-E [116]	MHOM RA [117]	Vis-QoL [118–124]	VFQ-UI [125–130]	Mac-New-7D [87]	ECO-HIS-4D [114]	PBI-WRQL [111]	PB-HIV [131]	CFQ-R-8 D [69]
Step 9. Select Rasch or IRT model and fit model																		
Rasch rating scale							X		X		X							
Rasch partial credit		X		X	X							X						
Unspecified polytymous Rasch	X							X		X			X	X	X	X	X	X
Graded response IRT			X			X												
Response level ordering[a]		X		X	X			X	X	X			X	X	X	X		X
Meaningfulness of merged levels	X	X								X								
Item parameters at logit 0	X				X			X				X						X
Global model fit – item-trait interaction[b]	X	X			X		X	X	X	X	X	X			X	X		X
Person separation index or person separation reliability[c]		X		X	X		X		X	X	X	X	X		X			X
Infit, outfit				X										X	X			
Item fit residuals[d]	X	X			X		X	X	X	X			X	X	X			X

Table 4 (continued)

Local dependence								X	X							X
Differential item functioning	X				X			X	X	X		X	X	X	X	X
Person fit residuals[e]					X		X									X
Targeting of scale to persons		X									X					
Unidimensionality – principal components analysis of residuals, independent t-statistics		X		X			X			X						
Step 10. Select items per dimension																
Item coverage across range of construct		X	X#										X		X	X
Absence of floor and ceiling effects		X										X	X	X		X
Percentage missing data												X				X
Correlation of item to dimension						X	X									X
Item importance and impact[e]									PCA							
Step 11. Model validation																
Expert validation													X			
Another data-set or split half	X	X			X					X	X					

Table 4 (continued)

Meaningful clinical rationale (face and content validity[e]		P, HCP	HCP	P	P, HCP	X
Alignment with parent psychometric instrument[C, E, P]	X					

[a] Adjacent response categories were merged if disordered

[b] Item-trait interaction fit statistics, n.s. chi-square, *p*-value > 0.01 after Bonferroni adjustment

[c] Person separation reliability > 0.7 for group use, or > 0.85 for individual use

[d] Exclude items where fit residuals > 2.5 or < -2.5 – and refit Rasch model

[e] Exclude persons where person fit residuals > 2.5 or < -2.5 and refit Rasch model

[f] P: patients, HCP: health care providers, R: researchers. #Fisher information for graded response model

[f] P: patients, HCP: health care providers, R: researchers, G: general public

into different levels of one item in the QLU-C10D [49]. DIF, or item bias, is when individuals with known attributes, such as gender or age with the same level of HRQoL respond differently [31, 65, 66, 75, 78, 83, 89, 103, 116, 127, 142]. Developers iteratively removed items with poor fit and the Rasch model was re-fit (Table 4).

Individuals with large person fit residuals (> 2.5 or < -2.5), representing outliers, were also removed, and the Rasch model was re-fit. Some developers evaluated how well the instrument targeted its respondents, with an expected person location of zero and a standard deviation of 1 [88, 127] (Table 4).

Lastly, some developers tested unidimensionality of the instrument by performing PCA of the item residuals after fitting the Rasch model. The associations between item residuals should be random. The developers of the DUI assessed the percentage of variance attributable to the Rasch factor, and the first residual factor to assess unidimensionality [65, 78]. Next, independent t-tests of person score residuals of items that loaded positively (> 0.30) or negatively (< -0.30) were sometimes performed. If the items in the instrument are strictly unidimensional, the percentage of significant tests should be < 5% (POS-E, P-PBMSI, DEMQOL-U, BUI) [65, 70, 103, 116]. This also can be expressed as a confidence interval for a binomial test of proportions for the significant tests (CORE-6D) [88] (Table 4).

Developers of the ReQoL and the NQU scoring system fit the GRM, an IRT model. The model fit of the ReQoL was evaluated with the sum-score-based item fit statistic (S- χ^2) [145]. The item information function was calculated to identify the score range where each item covered the most information, and the higher the discrimination parameter, the more information it provides. Test information of the total item pool was calculated, and the range where measurement precision > 0.9 was calculated [101, 145] (Table 4).

Step 10. Select items per dimension The purpose of selecting a small number of items per dimension is so that the health states from the eventual preference-based instrument are amenable to valuation [136]. This step overlaps with Brazier's stage II [5]. Developers used clinimetric and psychometric criteria to select items whether fitting a Rasch or IRT model. If items fit the Rasch model, most developers selected one item per dimension based on Rasch analysis criteria, conventional psychometrics, and item importance. Developers of the DMD-QoL-8D selected two items for each underlying factor [83]. Representative items for the dimension spanned a range of condition severity (AQL-5D, MSIS-8D, DMD-QoL-8D) [27, 33, 83] (Table 4). Developers retained items with a high correlation between the item and its dimension score (AQL-5D, DMD-QoL-8D) [33, 83], that could adequately discriminate (e.g., QLU-C10D and FACT-8D: early vs late stage cancer) [49, 62], or had high responsiveness (e.g., OAB-5D and FACT-8D: standardized response mean between baseline to specific time on treatment) [42, 62]. Conventional psychometric criteria were applied to exclude items with a high proportion of missing data (VFQ-UI, ABC-UI, DMD-QoL-8D) [31, 83, 126], or high floor and ceiling effects (VFQ-UI, CARIES-QC-U, DMD-QoL-8D) [83, 112, 126]. Some developers included item importance and impact ratings from experts to guide item selection (ABC-UI, QLU-C10D) [31, 49], or combinations of patient and health care provider perspectives (Table 4).

For the two instruments that used a graded response IRT model for item selection, developers chose items maximizing coverage of the construct, or selected two items per dimension for their item bank (Neuro-QoL) [101] (Table 4). Items with high Fisher information contribute to higher measurement precision (ReQoL) [94].

Step 11. Model validation The purpose of model validation is to evaluate whether the fitted model measures what it intended to measure [20]. Aligned with Brazier's stage IV [5] (Fig. 2), some developers validated the factor analysis or Rasch analysis using another dataset or a split half of the original dataset [27, 33, 43, 78, 84, 88, 116, 118]. Developers incorporated the perspectives of patients, clinicians, or researchers (e.g. importance ratings, interviews) to validate the meaningfulness of the resulting factors [45, 49, 62, 75, 78, 80, 84, 112]. Other developers checked that the resulting classification system had a dimensional structure aligned with the parent psychometric instrument [45, 49, 97] (Table 4).

Step 12. Evaluate measurement properties and interpretability The purpose of assessing measurement properties (reliability, validity, and responsiveness) of a novel instrument before it is used is so that we can be sure that it consistently measures what it is intended to measure, including changes in health [146]. Interpretability is being able to assign qualitative meaning to quantitative scores [146].

Developers evaluated construct validity [41, 63, 67, 72, 95, 103, 117, 127], criterion validity [49, 78, 103, 127] or reliability [34, 49, 63] (Table 5). Responsiveness between baseline and follow-up visits was used to select items (QLU-C10D, EORTC-8D, FACT-8D, AQL-5D) [33, 45, 49, 62]. For example, internal construct validity can be assessed by comparing response distributions with subscales of established instruments (DUI vs SF-12, W-BQ12, and DES) [78]. Criterion or convergent validity was evaluated relative to the parent instrument (IUI vs I-QOL)[40]. Other developers evaluated measurement

Table 5 Phase III (Step 12) Evaluate measurement properties and interpretability

	ABC-UI [31, 32]	AQL-5D [33–39]	IUI [40, 41]	OAB-5D [9, 36, 42–44]	EORTC-8D [7, 45–48]	QLU-C10D [49–60]	HAQ-PBM, QLQ-PBM, MSIS-PBM [61]	FACT-8D [62, 63]	CP-6D [66, 67]	DEMQOL-U & -Proxy-U [8, 71–74]	AD-5D [75–77]	CARIES-QC-U [112]	DUI [78, 79]	DHP-3D & 5D [80]	HASMID-8 & 10 [81, 82]	
Measurement properties																
Reliability																
Test re-test						+						+				
Inter-rater																
Construct validity																
Discriminant or known groups	+	+			+	+		±	+			+	+			
Convergent		+	+		+		HAQ-PBM (±) QLQ-PBM (+) MSIS-PBM (+)	+		+		+				
Criterion validity																
Predictive			+			+							+			
Responsiveness		+	+	+	+	+	QLQ-PBM (±)	+		+		+			+	
Interpretability																
Minimal important difference				X						X						
Item wording												X				
	NEWQoL-6D [84–86]	CORE-6D [88–90]	ReQoL-UI [91–94]	MobQoL-7D [95, 140]	MSIS-8D & -P [27, 97–100, 147]	Neuro-QoL derived NQU [101, 102]	P-PBMSI [103–106]	MF-8D [107]	WAITe [108, 109]	POS-E [116]	MHOMRA [117]	Vis-QoL [118–124]	VFQ-UI [125–130]	PB-WRQL [111]	PB-HIV [131]	ECOHIS-4D [148]

Table 5 (continued)

Measurement properties												
Reliability												
Test re-test		+			+	+						
Inter-rater					+							
Construct validity												
Discriminant or known groups		+	+	+	+		+	+	+	+	+	+
Convergent	+	+	+	+	+	+	+		+		+	
Criterion validity												
Predictive			+						±			
Responsiveness	±								+	+		+
Interpretability												
Minimal important difference												
Item wording											X	

+: Established measurement property, ±: Equivocal measurement property. X: Established properties after utilities were elicited (e.g., responsiveness of the DEMQOL-U vs EQ-5D-5L) [8] (Table 5).

The minimal important difference (MID) was assessed for the OAB-5D and compared with the EQ-5D-5L [9]. Both anchor and distribution-based methods were used to determine the MID of the DEMQOL-U [73] (Table 5).

Phase IV (Steps 13–15): Valuation and modeling of health state utilities

Table 6 outlines these steps, which are aligned with Brazier's stages V and VI [5]. The data required for phase IV are utility weights.

Step 13. Elicit heath state utility values The purpose of eliciting utility values is to develop a set of utility weights to assign to the health states derived from the instrument [149, 150]. Individuals eliciting utility weights were either patients, members of the general public, or carers. Twenty-five CSPBIs elicited utilities from the general public, the most common group, whereas 13 CSPBIs elicited utilities from patients. Patients produced significantly higher utility values than the general public when assessed for the same instrument (e.g., cognition in MS) (MSIS-8D) [98]. Health states must be selected for valuation, and the most common method was an orthogonal design in which each dimension level had an equal chance of combination with all other dimension levels in the instrument (15 instruments) (Table 6). Direct utilities were elicited using cardinal (e.g., TTO, SG), or ordinal (e.g., DCE) methods, most frequently using TTO (21 instruments) (Table 6).

Step 14. Model utility function Statistical models are used to generate a set of utility weights for responses from the CSPBI questionnaire [150]. Utilities of individual [31, 43, 80, 84], aggregate [31, 40, 43, 80, 84, 89], or re-scaled data were modeled [89]. Developers modeled utility values using additive or multiplicative models. Additive models were logit (conditional, mixed, or multinomial), ordinary least squares [31, 43, 80, 83, 84], or random effects models [61, 99]. Multiplicative models fit multiattribute utility functions [40, 79, 119]. Two groups of developers used pseudo preferences by regressing EQ-VAS on each dimension's response option to create preference weights (PB-WRQL, PB-HIV) [111, 131] (Table 6).

Step 15. Evaluate utility function Developers used various criteria to evaluate the utility function used to score the CSPBI. In our scoping review, the utility function was evaluated based on regression model coefficients for statistical significance [31, 43, 45, 80, 83, 89], and for consistency with the descriptive systems [31, 45, 80, 83]. For example, individuals with poor health were expected to have lower utility values than people with good health. Developers also evaluated the relationship between observed and model-derived utility values [31, 40, 80], or compared model-derived values with EQ-5D utilities [61]. Predictive validity of the algorithm compared estimated and observed utilities using sum of total differences, mean of differences, mean of absolute differences, overall standard deviation of differences, and intraclass correlation coefficient (ICC) [40]. Developers then assessed model fit using root mean square error (RMSE) [31, 80, 84, 89], mean absolute error (MAE) [31, 43, 45, 80, 84], Akaike's Information Criterion (AIC), and Bayesian Information Criterion (BIC) [31, 49], where smaller values indicate better fit. Model fit also was assessed with adjusted R^2 which is the proportion of variation explained by the model, where values approaching 1 are better [40, 43, 89, 125] (Table 6).

Figure 2 shows our 15-step framework with Brazier's corresponding stages.

Discussion

This scoping review produced a framework with 15 key steps that outline the phases of developing CSPBIs from the development of a conceptual framework to evaluating the utility function. This framework overlaps with the steps or stages from existing frameworks from psychometrics [16], and factor analysis [29, 151], and augments Brazier's six stages of CSPBI development [5]. Brazier's stages begin at our step 4 with establishing dimensionality of a pre-existing non-preference-based instrument. We added steps 1–3, required when developing any instrument de novo, coinciding with psychometric item development.

Our framework is novel by connecting the steps of initial stages of psychometric item development (phase I) established by Guyatt et al. (1986)'s seven stages of questionnaire design [16], with the steps of preference-instrument development. Our framework steps, excluding step 1 (a priori framework), are found within Guyatt's stages [16], but in a different order due to their emphasis on judgemental approaches in creating a psychometric questionnaire vs our focus on quantitative approaches to developing a preference-based questionnaire. We have noted that some steps in Phase II are desirable when performing EFA or PCA, but they are not required. In circumstances where data availability is limited, model validation using a novel dataset may not be possible.

Through comparing our approach with O'Brien [28], and Norman and Streiner [29], our framework generalizes those authors' approaches that are common amongst factor analysis with and without an a priori factor structure.

Deductive and inductive methods could be combined to reduce initial items. Deductive methods include

Table 6 Phase IV (Step 13–15) Value, model, and evaluate health state utilities

	ABC-UI [31, 32]	AQL-5D [33–39]	IUI [40, 41]	OAB-5D [9, 36, 42–44]	EORTC-8D [7, 45–48]	QLU-C10D [49–58]	HAQ-PBM, QLQ-PBM, MSIS-PBM [61]	FACT-8D [62, 63]	DEMQOL-U & -Proxy-U [8, 71–74]	AD-5D [75–77]	CARIES-QC-U [112, 113]	DUI [78, 79]	DHP-3D & 5D [80]	HASMID-8 & 10 [81, 82]	DMD-QoL-8D [83]	CP-6D [68]	ECHOHIS-4D [115]
13. Elicit health state utility values																	
Whose utilities																	
Patients										X		X		X			
General public	X	X	X	X	X	X	X		X	X	X		X	X	X	X	X
Carers									X[+++]	X							
Valuation method																	
TTO[a]	X	X	X	X	X		X		X				X				
DCE[b] *or DCE-TTO*[c]						X[++]		X		X	X			X	X	X	X
VAS[d]			X									X					
EQ-VAS																	
BWS[e]										X	X						
RS[f]																	
SG[g]												X					
Method of selecting health states																	
Orthogonal or balanced design	X	X		X	X	X[+]	X		X	X			X				
Single and multi-attribute health states to enable modeling			X									X					
D-efficiency										X	X			X		X	X
C-efficiency								X							X		
Corner states			X									X					
Rasch vignette																	
Own health state		X[@]															

Table 6 (continued)

Intermediate and anchor states		X															X
Naming health state																	
14. Model utility values																	
Individual level data	X	X		X	X	X	X		X	X	X		X	X	X	X	X
Aggregate (mean) data	X	X	X	X	X	X		X				X	X				
Functional form																	
Additive function		X				X	X	X						X		X	X
Multiplicative function		X^	X									X					
Model type																	
Conditional logit						X$^+$		X		X	X			X	X	X	X
Mixed logit						X$^+$		X								X	
Multinomial logit										X							
Ordinary least squares	X	X		X	X				X				X				
Tobit																	
Random effects							X										
Multiattribute utility function			X									X					
Estimation method																	
Maximum likelihood estimation	X	X			X												

Table 6 (continued)

Expected a posteriori				X													
15. Evaluate utility function																	
*Regression model coefficients**	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
*Consistency of coefficients with descriptive system***	X	X			X	X	X	X	X	X		X	X	X	X	X	X
Fit statistics																	
RMSE[§]	X								X				X				
MAE[§§]	X	X	X	X	X		X		X				X				
AIC[§§§]	X					X^{+}		X		X						X	X
BIC[§§§§]	X					X^{+}		X		X						X	X
R^2 / Adjusted R^2		X	X	X					X		X	X		$X^{\wedge\wedge}$	$X^{\wedge\wedge}$		
	WAITe [110]	NEWQoL-6D [84–86]	CORE-6D [88–90]	ReQoL-UI [91–94]	MSIS-8D & -P [27, 97–100]	Neuro-QoL derived NQU [101, 102]	P-PBMSI [103–106]	MF-8D [107]	MHOM RA [117]	Vis-QoL [118–124]	VFQ-UI [125–130]	PBI-WRQL [111]	PB-WRQL [111]	PB-HIV [131]	CFQ-R-8D [69]	MobQoL-7D [96]	
13. Elicit health state utility values																	
Whose utilities																	
Patients					X	X	X		X	X	X	X	X	X		X	
General public	X	X	X	X	X	X		X			X				X	X	
Carers																	
Valuation method																	
TTO[a]	X	X	X	X	X			X	X	X	X				X		
DCE[b] *or DCE-TTO*[c]	X																
VAS[d]												X				X	
EQ-VAS													X	X			
BWS[e]			X														
RS[f]							X			X							
SG[g]						X	X										

Table 6 (continued)

Method of selecting health states																
Orthogonal or balanced design		X						X								
Single and multi-attribute health states to enable modeling						X	X			X						
D-efficiency																
C-efficiency																
Corner states						X	X			X						
Rasch vignette			X	X												
Own health state													X	X		X
Intermediate and anchor states	X															X
Naming health state					X											
14. Model utility values																
Individual level data	X	X		X								X			X	X
Aggregate (mean) data		X	X			X	X			X						
Functional form																
Additive function		X	X		X	X			X		X	X			X	
Multiplicative function			X			X	X			X						
Logistic function													X	X		
Conditional logit																
Mixed logit																
Multinomial logit																

Table 6 (continued)

Ordinary least squares		X		X				X			X					
Tobit															X	
Random effects					X										X	
Multiattribute utility function										X						
Estimation method																
Maximum likelihood estimation				X				X								
Expected a posteriori																
15. Evaluate utility function																
*Regression model coefficients**	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X
*Consistency of coefficients with descriptive system***		X	X	X	X	X	X	X			X	X	X		X	
Fit statistics																
RMSE§		X	X	X	X						X					
MAE§§		X		X	X			X							X	
AIC§§§				X											X	
BIC§§§§				X											X	
R^2 / Adjusted R^2			X		X						X	X				

+EORTC-derived QLQ-C10 utility weights were elicited in different countries (Australia, UK, Canada, France, Germany, Netherlands, US) using similar methods.++EORTC-QLU C10D DCE-$_{TTO}$ valuation task included 11 attributes – the 10 dimensions of the QLU C10D and 1 attribute of time.+++Carers valued the DEMQOL Proxy-U

a: time trade off (TTO), b: discrete choice experiment (DCE), c: discrete choice experiment – time trade off (DCE-TTO), d: visual analogue scale (VAS), e: best worst scaling (BWS), f: rating scale (RS), g: standard gamble (SG)

@Respondents valued their own health state if they had asthma

[*] Regression model coefficient significance. **Consistency of coefficients with descriptive system parameters i.e., worse health should produce lower utilities. §root mean square error (RMSE), §§: mean absolute error (MAE), §§§: Akaike's information criterion (AIC), §§§§: Bayesian Information Criterion (BIC)

^Linear parametric models were fitted by Yang, and a multiplicative Bayesian models were fitted by Kharroubi.^^Pseudo R^2

selecting an a priori framework, using a Rasch or IRT model in item reduction, or hypothesis testing using CFA. Inductive methods include generating items from experts and performing EFA or PCA. Some developers used the results of EFA to inform hypothesis testing on another dataset to fit CFA models [62, 94].

Structural independence, where every health state defined by the set of attributes and levels is possible, could be better emphasized in the development of CSPBI, since it is a defining feature of preference-based instruments [152]. Structural independence can be inferred from factor analysis or evaluated using methods such as Rasch analysis [5], k-means cluster analysis [153], or pairwise independence of attributes [154, 155]. Only one instrument developer explicitly mentioned evaluating structural independence [45].

Surprisingly, few developers used CFA, even though most instruments were developed from existing psychometric instruments, when an a priori dimensional structure could be tested. When evaluating factor loadings, developers did not explicitly state the need to have 2 or 3 items per factor, or that a key objective in EFA is to fit the most parsimonious factor structure [29].

While 39 of 41 instruments used Rasch analysis, fewer than half of CSPBI developers explicitly described using psychometric and Rasch criteria in item selection (step 10), a critical step in this framework.

Peasgood et al. [136] described additional item selection criteria which are being applied in developing the novel generic preference-based instrument, the EQ-HWB (health and well-being) [156]. Some of these criteria overlap with the concept of sensibility [157] and coverage of the full range of the domain in our item selection step. Peasgood et al. also highlight ensuring measurement of current HRQoL so that items can be used in comparisons between and within people, and ensuring that the items are suitable for valuation [136].

The utility elicitation method, respondent type (general public vs patients), and the functional form likely affected the derived utility values but these were frequently not acknowledged and could be further studied [101, 158].

Limitations of this scoping review were not including a critical appraisal of included articles and only included CSPBIs in which Rasch analysis or IRT analysis were used in the steps of their development.

Conclusions

This study fills a gap in the methodological literature by providing a comprehensive framework to guide the development of preference-based instruments de novo, adding to quality assessment criteria of patient-reported outcomes such as the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) [146, 159]. Rasch and IRT methods improved item selection and the overall robustness of the resulting instruments with potential for item banking and computerized adaptive testing [101, 158]. This study will help guide the rigorous development of CSPBIs, to better measure patient preferences for clinical decision-making and cost-effectiveness analyses.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12955-024-02253-y.

Supplementary Material 1.

Acknowledgements
We thank Doyoung Kim for assistance with calculating reliability for this manuscript.

Authors' contributions
TCOT conceptualized and designed this scoping review; co-designed the search strategy; performed most current literature searches; co-conducted title and abstract screening, and full text screening; performed data abstraction, data analysis and interpretation; wrote the first draft of the manuscript and critically-revised this manuscript. SCT was the secondary reviewer for screening and data abstraction. JMB co-designed the search strategy and performed initial literature searches. NM, and MET provided guidance on conceptualizing the scoping review, and interpretation of the data. KEB assisted in data analysis, data interpretation, and manuscript revisions. MDK and AMD provided guidance on conceptualizing the scoping review and framework, summarizing, analyzing, and interpreting the data. All authors, except for MDK, critically reviewed the manuscript for publication and have approved the submitted version. All are accountable for the accuracy and integrity of the work.

Authors' information
See page 1.

Funding
The authors declare that no funds, grants, or other support were received for the study or preparation of this manuscript.

Availability of data and materials
Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1] Toronto Health Economics and Technology Assessment (THETA) Collaborative, University Health Network, Toronto, ON, Canada. [2] Canadian Centre for Applied Research in Cancer Control, Toronto, ON, Canada. [3] Child Health and Evaluative Sciences, Hospital for Sick Children, Toronto, ON, Canada. [4] Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. [5] Centro Hospitalar Universitário Lisboa Norte, Lisboa, Portugal. [6] Children's Hospital of Eastern Ontario, Ottawa, ON, Canada. [7] Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, ON, Canada.

Received: 7 May 2023 Accepted: 22 April 2024
Published online: 14 May 2024

References

1. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population survey. University of York. 1995.
2. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095–108.
3. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation: Oxford University Press; 2017.
4. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 4th ed. Oxford: Oxford University Press; 2015.
5. Brazier J, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, 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 Technology Assessment. 2012;16(32):https://doi.org/10.3310/hta16320.
6. Rowen D, Brazier J, Ara R, Azzabi ZI. The role of condition-specific preference-based measures in health technology assessment. Pharmacoeconomics. 2017;35(Suppl 1):33–41.
7. Lorgelly PK, Doble B, Rowen D, Brazier J. Cancer i. Condition-specific or generic preference-based measures in oncology? A comparison of the EORTC-8D and the EQ-5D-3L. Qual Life Res. 2017;26(5):1163–76.
8. Ratcliffe J, Flint T, Easton T, Killington M, Cameron I, Davies O, et al. An empirical comparison of the EQ-5D-5L, DEMQOL-U and DEMQOL-Proxy-U in a post-hospitalisation population of frail older people living in residential aged care. Appl Health Econ Health Policy. 2017;15(3):399–412.
9. Herdman M, Nazir J, Hakimi Z, Siddiqui E, Huang M, Pavesi M, et al. Assessing preference-based outcome measures for overactive bladder: an evaluation of patient-reported outcome data from the BESIDE clinical trial. The patient. 2017.
10. Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. Med Decis Making. 2008;28(1):66–89.
11. Petrillo J, Cairns J. Converting condition-specific measures into preference-based outcomes for use in economic evaluation. Expert Rev Pharmacoecon Outcomes Res. 2008;8(5):453–61.
12. Lin FJ, Longworth L, Pickard AS. Evaluation of content on EQ-5D as compared to disease-specific utility measures. Qual Life Res. 2013;22(4):853–74.
13. 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(2):161–83.
14. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. J Clin Oncol. 1993;11(3):570–9.
15. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, 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(5):365–76.
16. Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. CMAJ. 1986;134(8):889–95.
17. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual Life Res. 2007;16(Suppl 1):5–18.
18. Brazier JE, Rowen D, Mavranezouli I, Tsuchiya A, Young T, Yang Y, 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 Technology Assessment (Winchester, England). 2012;16(32):1–114.
19. 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(8):1358–62.
20. Clark LA, Watson D. Constructing validity: basic issues in objective scale development. Psychol Assess. 1995;7(3):309–19.
21. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc. 2015;13(3):141–6.
22. Peters M, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: Scoping Reviews (2020 version). In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis: JBI; 2020.
23. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med. 2018;169(7):467–73.
24. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol. 2016;75:40–6.
25. Covidence Melbourne, Victoria 3000, Australia2021 [Available from: https://www.covidence.org/.
26. De Vet H, Terwee C, Mokkink L, Knol D. Measurement in Medicine: a practical guide. Cambridge: Cambridge University Press; 2011.
27. 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 in Health. 2015;18(8):1016–24.
28. O'Brien K. Factor analysis: an overview in the field of measurement. Physiother Can. 2007;59:142–55.
29. Norman G, Streiner D. Chapter 19 Principal Components and Factor Analysis: Fooling Around with Factors. 3rd Edition Biostatistics: the Bare Essentials. Hamilton, ON: B.C. Decker, Inc.; 2008. p. 194–209.
30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.
31. Kerr C, Breheny K, Lloyd A, Brazier J, Bailey DB Jr, Berry-Kravis E, et al. Developing a utility index for the Aberrant Behavior Checklist (ABC-C) for fragile X syndrome. Qual Life Res. 2015;24(2):305–14.
32. Raspa M, Sacco P, Candrilli SD, Bishop E, Petrillo J. Validity of a condition specific outcome measure for fragile X syndrome: the Aberrant Behaviour Checklist-utility index. J Intellect Disabil Res. 2016;60(9):844–55.
33. 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.
34. Crossman-Barnes CJ, Sach T, Wilson A, Barton G. The construct validity and responsiveness of the EQ-5D-5L, AQL-5D and a bespoke TTO in acute asthmatics. Qual Life Res. 2020;29(3):619–27.
35. Yang Y, Brazier JE, Tsuchiya A, Young TA. Estimating a preference-based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire. Med Decis Making. 2011;31(2):281–91.
36. Kharroubi SA, Brazier JE, Yang Y. Modeling a preference-based index for two condition-specific measures (asthma and overactive bladder) using a nonparametric Bayesian method. Value in Health. 2014;17(4):406–15.
37. Sullivan PW, Ghushchyan VH, Campbell JD, Globe G, Bender B, Magid DJ. Measurement of utility in asthma: evidence indicating that generic instruments may miss clinically important changes. Qual Life Res. 2016;25(12):3017–26.
38. Kontodimopoulos N, Stamatopoulou E, Brinia A, Talias MA, Ferreira LN. Are condition-specific utilities more valid than generic preference-based ones in asthma? Evidence from a study comparing EQ-5D-3L and SF-6D with AQL-5D. Expert Rev Pharmacoecon Outcomes Res. 2018;18(6):667–75.
39. Sadatsafavi M, McTaggart-Cowan H, Chen W, Mark FitzGerald J, Economic Burden of Asthma Study G. Quality of Life and Asthma Symptom Control: Room for Improvement in Care and Measurement. Value in Health. 2015;18(8):1043–9.
40. Cuervo J, Castejon 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.
41. Castejon 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.

42. 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(2):253–65.
43. Yang Y, Brazier J, Tsuchiya A, Coyne K. Estimating a preference-based single index from the Overactive Bladder Questionnaire. Value Health. 2009;12(1):159–66.
44. Desroziers K, Aballea S, Maman K, Nazir J, Odeyemi I, Hakimi Z. Estimating EQ-5D and OAB-5D health state utilities for patients with overactive bladder. Health Qual Life Outcomes. 2013;11:200.
45. Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, Velikova G. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Value in Health. 2011;14(5):721–31.
46. Rowen D, Young T, Brazier J, Gaugris S. Comparison of generic, condition-specific, and mapped health state utility values for multiple myeloma cancer. Value Health. 2012;15(8):1059–68.
47. Kularatna S, Whitty JA, Johnson NW, Jayasinghe R, Scuffham PA. Development of an EORTC-8D utility algorithm for Sri Lanka. Med Decis Making. 2015;35(3):361–70.
48. Kularatna S, Whitty JA, Johnson NW, Jayasinghe R, Scuffham PA. A comparison of health state utility values associated with oral potentially malignant disorders and oral cancer in Sri Lanka assessed using the EQ-5D-3 L and the EORTC-8D. Health Qual Life Outcomes. 2016;14:101.
49. King MT, Costa DSJ, Aaronson NK, Brazier JE, Cella DF, Fayers PM, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. Qual Life Res. 2016;25(3):625–36.
50. King MT, Viney R, Simon Pickard A, Rowen D, Aaronson NK, Brazier JE, et al. Australian Utility Weights for the EORTC QLU-C10D, a Multi-Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. Pharmacoeconomics. 2018;36(2):225–38.
51. Norman R, Viney R, Aaronson NK, Brazier JE, Cella D, Costa DS, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Qual Life Res. 2016;25(3):637–49.
52. Gamper E-M, Holzner B, King MT, Norman R, Viney R, Nerich V, Kemmler G. Test-retest reliability of discrete choice experiment for valuations of QLU-C10D health states. Value in Health. 2018;21(8):958–66.
53. Norman R, Mercieca-Bebber R, Rowen D, Brazier JE, Cella D, Pickard AS, et al. U.K. utility weights for the EORTC QLU-C10D. Health Econ. 2019;28(12):1385–401.
54. McTaggart-Cowan H, King MT, Norman R, Costa DSJ, Pickard AS, Regier DA, et al. The EORTC QLU-C10D: the Canadian valuation study and algorithm to derive cancer-specific utilities from the EORTC QLQ-C30. MDM Policy Pract. 2019;4(1):2381468319842532.
55. Kemmler G, Gamper E, Nerich V, Norman R, Viney R, Holzner B, et al. German value sets for the EORTC QLU-C10D, a cancer-specific utility instrument based on the EORTC QLQ-C30. Qual Life Res. 2019;28(12):3197–211.
56. Jansen F, Verdonck-de Leeuw IM, Gamper E, Norman R, Holzner B, King M, et al. Dutch utility weights for the EORTC cancer-specific utility instrument: the Dutch EORTC QLU-C10D. Qual Life Res. 2021;30(7):2009–19.
57. Nerich V, Gamper EM, Norman R, King M, Holzner B, Viney R, Kemmler G. French value-set of the QLU-C10D, a cancer-specific utility measure derived from the QLQ-C30. Appl Health Econ Health Policy. 2021;19(2):191–202.
58. Revicki DA, King MT, Viney R, Pickard AS, Mercieca-Bebber R, Shaw JW, et al. United States utility algorithm for the EORTC QLU-C10D, a multiattribute utility instrument based on a cancer-specific quality-of-life instrument. Med Decis Making. 2021;41(4):485–501.
59. Gamper EM, King MT, Norman R, Efficace F, Cottone F, Holzner B, et al. EORTC QLU-C10D value sets for Austria, Italy, and Poland. Qual Life Res. 2020;29(9):2485–95.
60. Pilz MJ, Nolte S, Liegl G, King M, Norman R, McTaggart-Cowan H, et al. The European organisation for research and treatment of cancer quality of life utility-core 10 dimensions: development and investigation of general population utility norms for Canada, France, Germany, Italy, Poland, and the United Kingdom. Value Health. 2023;26(5):760–7.
61. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA. Condition-specific preference-based measures: benefit or burden? Value Health. 2012;15(3):504–13.
62. King MT, Norman R, Mercieca-Bebber R, Costa DSJ, McTaggart-Cowan H, Peacock S, et al. 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(6):862–73.
63. Herdman M, Kerr C, Pavesi M, Garside J, Lloyd A, Cubi-Molla P, Devlin N. Testing the validity and responsiveness of a new cancer-specific health utility measure (FACT-8D) in relapsed/refractory mantle cell lymphoma, and comparison to EQ-5D-5L. J Patient Rep Outcomes. 2020;4(1):22.
64. Tsui TCO, Trudeau M, Mitsakakis N, Torres S, Bremner KE, Kim D, 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(2):e0262635.
65. Tsui TCO, Trudeau ME, Mitsakakis N, Krahn MD, Davis AM. Developing the breast utility instrument to measure health-related quality-of-life preferences in patients with breast cancer: selecting the item for each dimension. MDM Policy Pract. 2022;7(2):23814683221142268.
66. Bahrampour 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(18):2647–55.
67. Bahrampour M, Downes M, Scuffham PA, Byrnes J. Comparing multi-attribute utility instruments: CP-6D, a Cerebral palsy specific instrument, vs AQoL-4D. Expert Rev Pharmacoecon Outcomes Res. 2022;22(2):217–24.
68. Bahrampour M, Norman R, Byrnes J, Downes M, Scuffham PA. Utility values for the CP-6D, a cerebral palsy-specific multi-attribute utility instrument, using a discrete choice experiment. Patient. 2021;14(1):129–38.
69. Acaster S, Mukuria C, Rowen D, Brazier JE, Wainwright CE, Quon BS, et al. Development of the cystic fibrosis questionnaire-revised-8 dimensions: estimating utilities from the cystic fibrosis questionnaire-revised. Value Health. 2023;26(4):567–78.
70. Mulhern B, Smith SC, Rowen D, Brazier JE, Knapp M, Lamping DL, et al. Improving the measurement of QALYs in dementia: developing patient- and carer-reported health state classification systems using Rasch analysis. Value in Health. 2012;15(2):323–33.
71. Mulhern B, Rowen D, Brazier J, Smith S, Romeo R, Tait R, et al. Development of DEMQOL-U and DEMQOL-PROXY-U: generation of preference-based indices from DEMQOL and DEMQOL-PROXY for use in economic evaluation. Health Technology Assessment (Winchester, England). 2013;17(5):v-xv, 1–140.
72. Easton T, Milte R, Crotty M, Ratcliffe J. An empirical comparison of the measurement properties of the EQ-5D-5L, DEMQOL-U and DEMQOL-Proxy-U for older people in residential care. Qual Life Res. 2018;27(5):1283–94.
73. Lee EC, Wright J, Walters SJ, Cooper CL, Mountain GA. Estimating the minimum important difference in the DEMQOL instrument in people with dementia. Qual Life Res. 2021.
74. Rowen D, Mulhern B, Banerjee S, Hout BV, Young TA, Knapp M, et al. Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-Proxy. Value Health. 2012;15(2):346–56.
75. 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 and Quality of Life Outcomes. 2017;15 (1) (no pagination)(21).
76. Comans TA, Nguyen KH, Mulhern B, Corlis M, Li L, Welch A, et al. Developing a dementia-specific preference–based quality of life measure (AD-5D) in Australia: a valuation study protocol. BMJ Open. 2018;8(1):e018996.
77. Comans TA, Nguyen KH, Ratcliffe J, Rowen D, Mulhern B. Valuing the AD-5D dementia utility instrument: an estimation of a general population tariff. Pharmacoeconomics. 2020;38(8):871–81.
78. Sundaram M, Smith MJ, Revicki DA, Elswick B, Miller L-A. Rasch analysis informed the development of a classification system for a diabetes-specific preference-based measure of health. J Clin Epidemiol. 2009;62(8):845–56.
79. Sundaram M, Smith MJ, Revicki DA, Miller L-A, Madhavan S, Hobbs G. Estimation of a valuation function for a diabetes mellitus-specific

preference-based measure of health: the diabetes utility index. Pharmacoeconomics. 2010;28(3):201–16.

80. Mulhern B, Labeit A, Rowen D, Knowles E, Meadows K, Elliott J, Brazier J. Developing preference-based measures for diabetes: DHP-3D and DHP-5D. Diabet Med. 2017;34(9):1264–75.
81. Carlton J, Rowen D, Elliott J. Assessment of the psychometric properties and refinement of the Health and Self-Management in Diabetes Questionnaire (HASMID). Health Qual Life Outcomes. 2020;18(1):59.
82. Rowen D, Stevens K, Labeit A, Elliott J, Mulhern B, Carlton J, et al. Using a discrete-choice experiment involving cost to value a classification system measuring the quality-of-life impact of self-management for diabetes. Value Health. 2018;21(1):69–77.
83. Rowen D, Powell P, Mukuria C, Carlton J, Norman R, Brazier J. Deriving a preference-based measure for people with duchenne muscular dystrophy from the DMD-QoL. Value in Health. 2021;24(10):1499–510.
84. Mulhern B, Rowen D, Jacoby A, Marson T, Snape D, Hughes D, et al. The development of a QALY measure for epilepsy: NEWQOL-6D. Epilepsy Behav. 2012;24(1):36–43.
85. Mulhern B, Pink J, Rowen D, Borghs S, Butt T, Hughes D, et al. Comparing generic and condition-specific preference-based measures in epilepsy: EQ-5D-3L and NEWQOL-6D. Value Health. 2017;20(4):687–93.
86. Mulhern B, Rowen D, Snape D, Jacoby A, Marson T, Hughes D, et al. Valuations of epilepsy-specific health states: a comparison of patients with epilepsy and the general population. Epilepsy Behav. 2014;36:12–7.
87. Kularatna S, Rowen D, Mukuria C, McPhail S, Chen G, Mulhern B, et al. Development of a preference-based heart disease-specific health state classification system using MacNew heart disease-related quality of life instrument. Qual Life Res. 2021.
88. Mavranezouli 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(3):321–33.
89. Mavranezouli 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(3):381–95.
90. Wickramasekera N, Tubeuf S. Measuring quality of life for people with common mental health problems. J Ment Health. 2020:1–8.
91. Connell J, Brazier J, O'Cathain A, Lloyd-Jones M, Paisley S. Quality of life of people with mental health problems: a synthesis of qualitative research. Health Qual Life Outcomes. 2012;10:138.
92. Connell J, O'Cathain A, Brazier J. Measuring quality of life in mental health: are we asking the right questions? Soc Sci Med. 2014;120:12–20.
93. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technology Assessment (Winchester, England). 2014;18(34):vii-viii, xiii-xxv, 1–188.
94. Keetharuth AD, Rowen D, Bjorner 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(2):281–90.
95. Bray N, Tudor Edwards R. Preference-based measurement of mobility-related quality of life: developing the MobQoL-7D health state classification system. Disabil Rehabil. 2020:1–15.
96. Bray N, Tudor Edwards R, Schneider P. Development of a value-based scoring system for the MobQoL-7D: a novel tool for measuring quality-adjusted life years in the context of mobility impairment. Disabil Rehabil. 2024:1–10.
97. Goodwin E, Green C, Spencer A. Estimating a preference-based index for an eight-dimensional health state classification system for multiple sclerosis. Value in Health. 2015;18(8):1025–36.
98. Goodwin E, Green C, Hawton A. What difference does it make? A comparison of health state preferences elicited from the general population and from people with multiple sclerosis. Value Health. 2020;23(2):242–50.
99. Goodwin E, Green C, Hawton A. Health state values derived from people with multiple sclerosis for a condition-specific preference-based measure: multiple sclerosis impact scale-eight dimensions-patient version (MSIS-8D-P). Value Health. 2018;21(11):1338–45.
100. Green C, Goodwin E, Hawton A. "Naming and Framing": the impact of labeling on health state values for multiple sclerosis. Med Decis Making. 2017;37(6):703–14.
101. Matza LS, Phillips G, Dewitt B, Stewart KD, Cella D, Feeny D, et al. A Scoring algorithm for deriving utility values from the Neuro-QoL for patients with multiple sclerosis. Med Decis Making. 2020;40(7):897–911.
102. Matza LS, Stewart KD, Phillips G, Delio P, Naismith RT. Development of a brief clinician-reported outcome measure of multiple sclerosis signs and symptoms: The Clinician Rating of Multiple Sclerosis (CRoMS). Mult Scler Relat Disord. 2019;35:253–61.
103. 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(5):1445–57.
104. Kuspinar A, Pickard S, Mayo NE. Developing a valuation function for the preference-based multiple sclerosis index: comparison of standard gamble and rating scale. PLoS ONE. 2016;11(4):e0151905.
105. Kuspinar A, Bouchard V, Moriello C, Mayo NE. Development of a bilingual MS-specific health classification system: the preference-based multiple sclerosis index. Int J MS Care. 2016;18(2):63–70.
106. Kuspinar A, Mayo NE. Validation of the preference-based multiple sclerosis index. Mult Scler. 2019;25(11):1496–505.
107. 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 in Health. 2015;18(6):846–55.
108. 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(8):2025–39.
109. Oluboyede Y, Tomos R. Measuring weight-specific quality of life in adolescents: an examination of the concurrent validity and test-retest reliability of the WAItE. Value Health. 2019;22(3):348–54.
110. Robinson T, Hill S, Orozco-Leal G, Kernohan A, King W, Oluboyede Y. Valuing selected WAItE health states using the Time Trade-Off methodology: findings from an online interviewer-assisted remote survey. J Patient Rep Outcomes. 2024;8(1):6.
111. Moga AM, Twells LK, Mayo NE. The prototype of a preference-based index of weight-related quality of life: demonstrating the possibilities. Qual Life Res. 2022;24:24.
112. 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(1):105.
113. Rogers HJ, Sagabiel J, Marshman Z, Rodd HD, Rowen D. Adolescent valuation of CARIES-QC-U: a child-centred preference-based measure of dental caries. Health Qual Life Outcomes. 2022;20(1):18.
114. Kularatna S, Amilani U, Senanayake S, Tonmukayakul U, Jamieson L, Arrow P. Developing an early childhood oral health impact-specific health-state classification system for a new preference-based instrument, the ECOHIS-4D. Community Dentistry and Oral Epidemiology. 2021;18.
115. Hettiarachchi RM, Arrow P, Senanayake S, Carter H, Brain D, Norman R, et al. Developing an Australian utility value set for the Early Childhood Oral Health Impact Scale-4D (ECOHIS-4D) using a discrete choice experiment. Eur J Health Econ. 2023;24(8):1285–96.
116. Dzingina M, Higginson IJ, McCrone P, Murtagh FEM. Development of a patient-reported palliative care-specific health classification system: the POS-E. Patient Patient-Centered Outcomes Res. 2017;10(3):353–65.
117. Chiou C-F, Suarez-Almazor ME, Sherbourne CD, Chang C-H, Reyes C, Dylan M, et al. Development and validation of a preference weight multiattribute health outcome measure for rheumatoid arthritis. J Rheumatol. 2006;33(12):2409–11.
118. Misajon R, Hawthorne G, Richardson J, Barton J, Peacock S, Iezzi A, Keeffe J. Vision and quality of life: the development of a utility measure. Invest Ophthalmol Vis Sci. 2005;46(11):4007–15.
119. Peacock S, Misajon R, Iezzi A, Richardson J, Hawthorne G, Keeffe J. Vision and quality of life: development of methods for the VisQoL vision-related utility instrument. Ophthalmic Epidemiol. 2008;15(4):218–23.
120. Finger RP, Hoffmann AE, Fenwick EK, Wolf A, Kampik A, Kernt M, et al. Patients' preferences in treatment for neovascular age-related macular degeneration in clinical routine.[Erratum appears in Br J Ophthalmol. 2012 Sep;96(9):1276 Note: Finger, Robert [corrected to Finger, Robert P]]. British Journal of Ophthalmology. 2012;96(7):997–1002.

121. Gothwal VK, Bagga DK. Utility values in the visually impaired: comparing time-trade off and VisQoL. Optom Vis Sci. 2013;90(8):843–54.
122. Sahebjada S, Fenwick EK, Xie J, Snibson GR, Daniell MD, Baird PN. Impact of keratoconus in the better eye and the worse eye on vision-related quality of life. Invest Ophthalmol Vis Sci. 2014;55(1):412–6.
123. Duncan JL, Richards TP, Arditi A, da Cruz L, Dagnelie G, Dorn JD, et al. Improvements in vision-related quality of life in blind patients implanted with the Argus II Epiretinal Prosthesis. Clin Exp Optom. 2017;100(2):144–50.
124. Fenwick EK, Xie J, Pesudovs K, Ratcliffe J, Chiang PP, Finger RP, Lamoureux EL. Assessing disutility associated with diabetic retinopathy, diabetic macular oedema and associated visual impairment using the Vision and Quality of Life Index. Clin Exp Optom. 2012;95(3):362–70.
125. Kowalski JW, Rentz AM, Walt JG, Lloyd A, Lee J, Young TA, 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(2):323–34.
126. Rentz AM, Kowalski JW, Walt JG, Hays RD, Brazier JE, Yu R, et al. Development of a preference-based index from the national eye institute visual function questionnaire-25. JAMA Ophthalmology. 2014;132(3):310–8.
127. Goh RL, Fenwick E, Skalicky SE. The visual function questionnaire: utility index: does it measure glaucoma-related preference-based status? J Glaucoma. 2016;25(10):822–9.
128. Naik RK, Gries KS, Rentz AM, Kowalski JW, Revicki DA. Psychometric evaluation of the national eye institute visual function questionnaire and visual function questionnaire utility index in patients with non-infectious intermediate and posterior uveitis. Qual Life Res. 2013;22(10):2801–8.
129. Brazier J, Muston D, Konwea H, Power GS, Barzey V, Lloyd A, et al. Evaluating the relationship between visual acuity and utilities in patients with diabetic macular edema enrolled in intravitreal aflibercept studies. Invest Ophthalmol Vis Sci. 2017;58(11):4818–25.
130. Rojananuangnit K, Sudjinda N. Utility index and patient-reported outcome measures in glaucomatous patients comparing with normal participants. Clin Ophthalmol. 2021;15:835–43.
131. Mate KKV, Lebouche B, Brouillette MJ, Fellows LK, Mayo NE. Development of a prototype for a bilingual patient-reported outcome measure of the important health aspects of quality of life in people living with HIV: the Preference Based HIV Index (PB-HIV). J Person Med. 2022;12(12):2080.
132. Ware J, Brook R, Davis-Avery A, Williams K, Stewart A, Rogers W, et al. Conceptualization and Measurement of Health for Adults in the Health Insurance Study. Vol 1: Model of Health and Methodology. Santa Monica, CA: RAND Corporation; R-1987/1980.
133. Ware J. Methodology in behavioral and psychosocial cancer research. Cancer. 1984;53:2316–26.
134. Powell PA, Carlton J, Rowen D, Chandler F, Guglieri M, Brazier JE. Development of a new quality of life measure for duchenne muscular dystrophy using mixed methods: the DMD-QoL. Neurology. 2021;96(19):e2438–50.
135. Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, Audin K. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. Br J Psychiatry. 2002;180:51–60.
136. Peasgood T, Mukuria C, Carlton J, Connell J, Brazier J. Criteria for item selection for a preference-based measure for use in economic evaluation. Qual Life Res. 2021;30(5):1425–32.
137. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care. 2002;40(2):113–28.
138. Powell PA, Carlton J, Rowen D, Brazier JE. Producing a preference-based quality of life measure for people with Duchenne muscular dystrophy: a mixed-methods study protocol. BMJ Open. 2019;9(3):e023685.
139. Powell PA, Carlton J, Woods HB, Mazzone P. Measuring quality of life in Duchenne muscular dystrophy: a systematic review of the content and structural validity of commonly used instruments. Health Qual Life Outcomes. 2020;18(1):263.
140. Bray N, Spencer LH, Tuersley L, Edwards RT. Development of the MobQoL patient reported outcome measure for mobility-related quality of life. Disabil Rehabil. 2020:1–10.
141. Costa DS, Aaronson NK, Fayers PM, Grimison PS, Janda M, Pallant JF, et al. 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. Patient Related Outcome Measures. 2014;5:119–29.
142. Bahrampour 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. 2020:1–9.
143. Andrich D, Marais I. A Course in Rasch Measurement Theory. Crawley: Springer; 2019.
144. Maydeu-Olivares A, Montano R. How should we assess the fit of Rasch-type models? Approximating the power of goodness-of-fit statistics in categorical data analysis. Psychometrika. 2013;78:116–33.
145. Keetharuth AD, Bjorner JB, Barkham M, Browne J, Croudace T, Brazier J. An item response theory analysis of an item pool for the recovering quality of life (ReQoL) measure. Qual Life Res. 2021;30(1):267–76.
146. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–45.
147. Heather A, Goodwin E, Green C, Morrish N, Ukoumunne OC, Middleton RM, Hawton A. Multiple sclerosis health-related quality of life utility values from the UK MS register. Mult Scler J Exp Transl Clin. 2023;9(2):20552173231178440.
148. Weerasuriya SR, Hettiarachchi RM, Kularatna S, Rohde A, Arrow P, Jamieson L, et al. Comparison of the Early Childhood Oral Health Impact Scale (ECOHIS-4D) and Child Health Utility Index (CHU-9D) in children with oral diseases. Community Dent Oral Epidemiol. 2023.
149. Meltzer DO, Basu A, Sculpher MJ. Theoretical Foundations of Cost-Effectiveness Analysis in Health and Medicine. In: Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG, editors. Cost-Effectiveness in Health and Medicine. 2nd ed. United States of America: Oxford University Press; 2017. p. 496.
150. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada. CADTH; 2017 October 20,2016.
151. O' Brien K. Factor analysis: an overview in the field of measurement. Physiother Can. 2007;59:142–55.
152. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes. 2003;1:54.
153. Sugar CA, Sturm R, Lee TT, Sherbourne CD, Olshen RA, Wells KB, Lenert LA. Empirically defined health states for depression from the SF-12. Health Serv Res. 1998;33(4 Pt 1):911–28.
154. Dewitt B, Feeny D, Fischhoff B, Cella D, Hays RD, Hess R, et al. Estimation of a preference-based summary score for the patient-reported outcomes measurement information system: the PROMIS((R))-Preference (PROPr) Scoring System. Med Decis Making. 2018;38(6):683–98.
155. Hanmer J, Cella D, Feeny D, Fischhoff B, Hays RD, Hess R, et al. Selection of key health domains from PROMIS((R)) for a generic preference-based scoring system. Qual Life Res. 2017;26(12):3377–85.
156. Brazier J, Peasgood T, Mukuria C, Marten O, Kreimeier S, Luo N, et al. The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. Value in Health. 2022;25(4):482–91.
157. Feinstein A. The Theory and Evaluation of Sensibility. Clinimetrics. New Haven, U.S.A.: Yale University Press; 1987. p. pp. 141–65.
158. 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(3):38–47.
159. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet H, C,W,, Terwee CB. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) user manual. Amsterdam: COSMIN; 2018. p. 78.

Publisher's Note