CrossMark

ORIGINAL RESEARCH ARTICLE

Development of a Patient-Reported Palliative Care-Specific Health Classification System: The POS-E

Mendwas Dzingina[1] · Irene J. Higginson[1] · Paul McCrone[2] · Fliss E. M. Murtagh[1]

Abstract

Background Generic preference-based measures are commonly used to estimate quality-adjusted life-years (QALYs) to inform resource-allocation decisions. However, concerns have been raised that generic measures may be inappropriate in palliative care.

Objective Our objective was to derive a health-state classification system that is amenable to valuation from the ten-item Palliative Care Outcome Scale (POS), a widely used patient-reported outcome measure in palliative care.

Methods The dimensional structure of the original POS was assessed using factor analysis. Item performance was assessed, using Rasch analysis and psychometric criteria, to enable the selection of items that represent the dimensions covered by the POS. Data from six studies of patients receiving palliative care were combined ($N = 1011$) and randomly split into two halves for development and validation. Analysis was undertaken on the development data, and results were validated by repeating the analysis with the validation dataset.

Results Following Rasch and factor analyses, a classification system of seven items was derived. Each item had two to three levels. Rasch threshold map helped identify a set of 14 plausible health states that can be used for the valuation of the instrument to derive a preference-based index.

Conclusion Combining factor analysis and Rasch analysis with psychometric criteria provides a valid method of constructing a classification system for a palliative care-specific preference-based measure. The next stage is to obtain preference weights so the measure can be used in economic evaluations in palliative care.

Key Points for Decision Makers

We propose a new palliative care health-state classification system termed Palliative Care Outcome Scale (POS)-E.

POS-E classifies palliative care states as a combination of seven dimensions.

The dimensions are pain, other symptoms, anxiety, depression, family anxiety, feeling good about oneself and practical matters.

✉ Mendwas Dzingina
Mendwas.dzingina@kcl.ac.uk

1 Department of Palliative Care, Policy and Rehabilitation, Cicely Saunders Institute, King's College London, Denmark Hill, London SE5 9PJ, UK

2 King's Health Economics, King's College London, De Crespigny Park, Denmark Hill, London SE5 8AF, UK

1 Introduction

Economic evaluations are performed to inform the allocation of resources between competing healthcare interventions. A commonly used method is cost-utility analysis, which compares interventions in terms of their cost per quality-adjusted life-years (QALYs) gained. The QALY combines life expectancy (in years) and quality of life (QOL; expressed in the form of 'health state values') into a single metric based on people's preferences [1]. The QOL portion is estimated by assigning a numerical value to each health state experienced by a person on a scale ranging

△ Adis

from 1 (equivalent to full health) to 0 (dead) [2]. A common way of estimating health-state values is to use a 'generic' preference-based measure (PBM) such as the EuroQol five-dimensional questionnaire (EQ-5D) [3], Health Utilities Index Mark 3 (HUI3) [4], or Short-Form 6-Dimensions (SF-6D) [5]. Each generic PBM, e.g. EQ-5D, has a preference-based algorithm for assigning values to each health state. These preference weights are obtained by asking members of the general public to value the health states using a choice-based valuation technique such as standard gamble [6, 7] or time trade-off [6].

These generic PBMs are deemed appropriate for all patients, irrespective of their medical condition, because they concentrate on broad aspects of health-related QOL (HRQoL). However, debate has focussed on the degree to which the broad nature of these PBMs incorporates attributes of HRQoL that are particularly relevant to specific health conditions and health disciplines [8]. The estimation of QALYs in palliative care is one such case.

Palliative care is "the active holistic care of patients with advanced progressive disease, aimed at achieving the best possible QoL for patients and families, through the management of pain and other symptoms, as well as provision of spiritual, psychological and social support; which may be initiated early in the course of treatment along with other curative treatments" [9]. In the discipline of palliative care, there are concerns that generic PBMs do not incorporate many aspects of HRQoL important to patients receiving palliative care and rather are heavily focused on function (e.g. mobility, self-care and usual activities) [10–12]. This has led to proposals for the development of a condition-specific PBM (CSPBM) that would be appropriate for patients receiving palliative care [10, 13]. Furthermore, the likely dominant nature of palliative care needs in determining HRQoL arguably justifies the development and use of a CSPBM in palliative care. Presently, no such measure exists. The Palliative Care Outcome Scale (POS) has been suggested as suitable for this purpose [10]. The POS is a validated palliative care outcome measure [14] that has been used in many studies, including randomized controlled trials (RCTs) and observational studies, as well as for service evaluation [15–22]. Given the dearth of economic evaluations in palliative care [23], developing a CSPBM from a widely accepted and commonly used instrument such as the POS enables retrospective analysis of existing datasets and increases the likelihood that the measure will be used in future studies [24].

The process of developing a PBM from an existing condition-specific outcome measure involves three stages [8]. This paper reports on the first stage; the second and third stages will be addressed in a separate paper.

2 Methods

2.1 Design

This study was a secondary analysis of baseline data from several studies of patients receiving palliative care.

A health-state classification is a multidimensional framework that can be used to define health states. Such classifications define a set of health states by selecting one level from each dimension. For example, the EQ-5D has five dimensions, each comprising three levels of response, and defines a total of 243 states (3^5). This presents a more manageable number to value (and even then only a sample of states were directly valued). The POS has ten items, eight of which have five levels, and two items have three levels each. Given the number of items and their corresponding levels, the POS would define a practically unmanageable number of 3,515,625 health states ($5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 3 \times 3$). This would result in unreasonable cognitive demands on respondents to the valuation exercise required to estimate quality weights. Therefore, the first stage of deriving a health-state classification that is amenable to valuation from an existing measure involves using Rasch analysis to reduce the size of the existing measure while minimizing the loss of descriptive information [8]. This classification system would be designed to capture the range of palliative care-related problems that can occur with different diagnosis with minimal loss of information and the ability to use the responses from the original instrument to map onto it. Although some studies have derived and valued health-state classifications using standard methods (e.g. factorial and orthogonal block designs) that do not require a reduction in the size of the existing measure, such methods are inefficient because they treat items as independent (uncorrelated) statements and so are likely to result in deriving (and valuing) implausible health states. It is unlikely that the types of problems seen in palliative care are unrelated (as is implied in orthogonal and factorial designs). For example, it makes no sense to define a health state where a person feels 'good about themselves always' but also feels 'depressed always' as they are both likely to have the same primary cause. This approach of developing a health-state classification by using Rasch to reduce a larger instrument has been applied to numerous non-preference-based measures, including the SF-36 [25], SF-12 [26], menopausal health questionnaire [27], a preference-based measure for atopic dermatitis [28], King's Health Questionnaire [29], Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) [30] and European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire 30 (EORTC QLQ-C30) [31].

This study used a four-stage process as recommended by Brazier et al. [8] as follows:

1. Identify the most relevant dimensions of the POS for use in the POS-E, giving an initial descriptive system.
2. Identify item response levels that could be removed from the new descriptive system.
3. Identify item response levels that can be merged without loss of information.
4. Validate the new instrument by repeating steps 1–3 above in a separate dataset.

2.2 Datasets

We merged the following baseline POS data from six studies of patients receiving palliative care.

1. A cancer mortality follow-back survey ($N = 596$) from 2009 to 2010 in London (The QUALYCARE study) [32].
2. A study of Parkinson's disease (longitudinal community study of predictive factors; $N = 82$) [33].
3. An RCT on the effectiveness of an integrated palliative and respiratory care service for patients with advanced disease and refractory breathlessness in 2014 in the UK ($N = 105$) [12].
4. A longitudinal study on trajectories of illness of stage 5 chronic renal disease in the UK ($N = 74$) [34].
5. A cross-sectional study on symptom burden and palliative care needs in chronic obstructive pulmonary disease and cancer in Germany ($N = 109$) [15].
6. A randomised phase II trial of dignity therapy in the UK ($N = 45$) [35].

We then randomly split the data into a development dataset ($N = 504$) and a validation dataset ($N = 508$), providing suitable sample sizes for Rasch analysis. There is evidence that some Rasch fit statistics for polytomous instruments (e.g. POS) are sensitive to the sample size, and larger samples can have a higher chance of type 1 errors [36]. The development dataset was used to develop the health classification, and this was validated by repeating the analysis on the validation dataset. See the appendix (Table 7) for the descriptive statistics for each dataset. All datasets were anonymized prior to analysis.

2.3 The Palliative Care Outcome Scale (POS)

The ten-item POS is a short easy-to-use clinical outcome measure originally developed and validated in eight end-of-life and palliative care settings in the UK, including hospital, community, inpatient hospice, outpatient, day care and general practice [14, 37]. It was developed to measure domains that impact on the QOL of patients

receiving palliative care. The questionnaire consists of ten items, each item scored on a 5-point Likert scale ranging from 0 to 4, except items 9 and 10 ('time wasted' and 'practical matters'), both of which are scored on a 3-point scale (0, 2 and 4) as shown in the Electronic Supplementary Material (ESM) 1. The POS has been well validated and is widely used in clinical practice and research regionally and nationally in the UK to evaluate and improve the quality of care, and has been culturally adapted for use in 20 EU countries, Africa and other countries around the globe [15–22]. Two systematic reviews (in 2011 [39] and 2015 [38]) on the use of the POS found it was used in 78 published studies in both patients with and without cancer.

2.4 Analysis

The objective of the analysis was to derive a multi-dimensional health-state classification system amenable to valuation by reducing the number of items and item levels in the POS.

2.4.1 Step 1: Establishing Dimensions

Principal component analysis (PCA) was used to assess the dimensions of the POS. PCA is commonly used in the development of new instruments to provide early indications of possible dimensions before Rasch analysis is attempted [40]. First, the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy was used to assess the appropriateness of POS data for PCA (the KMO value should be >0.5 if the data are appropriate) [41]. In addition, Bartlett's test of sphericity was used to test whether the correlations between POS items were significant [42]. Significant factors (dimensions) were identified using Horn's parallel analysis [43] incorporated into an online facility by Watkins [44]. Next, the rotated factor matrices were examined to assess correlations of every item with each of the main factors of the instrument. We used both orthogonal and oblique rotation methods and compared the results of both, as recommended in the literature [45]. In all matrices, loadings with coefficients $\geq|0.400|$ were considered to reveal strong correlations between an item and a factor. Items loading on the same factor were considered to belong to the same underlying dimension captured by the POS.

2.4.2 Step 2: Eliminating Items Per Dimension

Rasch analysis was used to reduce the POS to a simpler descriptive health-state classification system by identifying POS items that did not fit the Rasch model and therefore were potentially unsuitable for inclusion in the

classification system. Rasch analysis is a mathematical technique used to convert categorical data to continuous data [46]. Rasch methods can be used to assess the extent to which individual items represent the underlying construct that an instrument intends to measure, thus enabling the assessment of the appropriateness of items for a classification system.

The following criteria were considered for item exclusion, in line with recommendations for multidimensional measures [8]:

- Item-level ordering (disordered thresholds): we examined threshold maps to identify items that had disordered thresholds. For instance, ordered thresholds indicate that a person with a high level of an attribute, such as pain, is more likely to endorse a high level on an item that measures pain than is a person with less pain. Disordered thresholds suggest that respondents are unable to differentiate between adjacent item categories [47]. In such instances, adjacent response categories were merged to obtain ordered thresholds. Items were excluded if their thresholds remained disordered despite merging of adjacent response categories. Furthermore, if the only way to obtain an ordered threshold for an item was by merging adjacent response categories in a way that did not make clinical sense, then such an item was eliminated. For example, it was deemed clinically meaningless to merge response categories 'moderately' and 'severely', as these indicate significantly different levels of severity.
- Rasch goodness of fit: following threshold re-ordering, overall and item-specific fit statistics were inspected to assess the extent to which the entire instrument, as well as individual items, fit the Rasch model. Items were excluded if fit residuals were >2.5 or less than −2.5 and/or chi-squared statistics were significant at the 0.001 level after Bonferroni adjustment [8].
- Differential-item functioning (DIF): items that demonstrate significant DIF are items with response patterns that vary according to specific patient factors such as diagnosis, age group, sex or ethnicity. Such items were excluded from further consideration because DIF can be a source of misfit in the Rasch model and because items forming a PBM should ideally express the same aspects of HRQoL across the whole patient population (and not distinguish significantly among subgroups with different baseline characteristics).

2.4.3 Step 3: Item Level Reduction

Rasch analysis can identify response levels that may be merged without losing descriptive information, offering further means of simplifying the classification system [8].

We identified potential item categories for merging by examining Rasch category probability curves and response frequencies. Visual inspection of respective category probability curves determined which adjacent response categories to merge. We also sought expert opinion about the clinical and psychometric meaningfulness of the merged item levels. These experts included a professor of psychology (Dr. R. Siegert, Auckland University of Technology, New Zealand) and two palliative care clinicians (Dr. P. Edmunds, King's College Hospital, London, and Dr. P. Kane, Beaumont Hospital, Dublin).

We also assessed the unidimensionality of the new classification system by using the test proposed by Smith [48], which involves conducting paired t tests of the final models. Unidimensionality is confirmed when ≤5% of the tests are significant at the $p < 0.05$ level [49]. We also examined the person separation index (PSI) to assess how efficiently the final set of items was able to separate those people measured. PSI values range from 0.0 to 1, with higher values indicating better separation and a more precise measure [49].

2.4.4 Step 4: Validation of Classification System

The health-state classification was validated by repeating steps 1–3 of the analysis using the validation data. We inspected the examining overall and item fit statistics, DIF, unidimensionality and item–response combinations.

RUMM2020 was used for all Rasch analysis and STATA version 12 for all other statistical analysis.

3 Results

3.1 Step 1: Factor Analysis

The KMO measure of sampling adequacy reached 0.79, suggesting that factoring of data was appropriate and meaningful. Bartlett's test of sphericity demonstrated the statistical significance of the findings ($p < 0.0001$). Although the analysis identified three factors with eigenvalues above 1, which explained 52% of the total variance (see Table 8 in the appendix for details), Horn's parallel analysis indicated two significant factors (Table 1). The scree plot (Fig. 1) appears to support a two-factor solution as the slope of the line flattens after the second factor.

In line with results of parallel analysis, a two-factor solution was extracted for rotation. Table 2 shows two rotated factors, one comprising six items (primarily about psychological and physical wellbeing) and the other comprising three items (two relating to the standard of care and one relating to psychological wellbeing). One item (time wasted) did not load above 0.40 on either of the two factors. Results were very similar between the two methods of

Table 1 Significant components of the Palliative Care Outcome Scale identified by principal component analysis (N = 504), and comparison of components with eigenvalues >1 with significant components identified by Horn's parallel analysis

Component	PCA: initial eigenvalues			Horn's parallel analysis: significant mean eigenvalues (SD)	
	Total	% of variance	Cumulative %		
1	**2.908**	29.080	32.807	**1.2609**	(0.0359)
2	**1.269**	12.693	41.773	**1.1833**	(0.0355)
3	1.013	10.128	51.901	**1.1134**	(0.0242)

Bold formatting indicates the significant eigenvalue levels identified using each approach

PCA principal component analysis, *SD* standard deviation
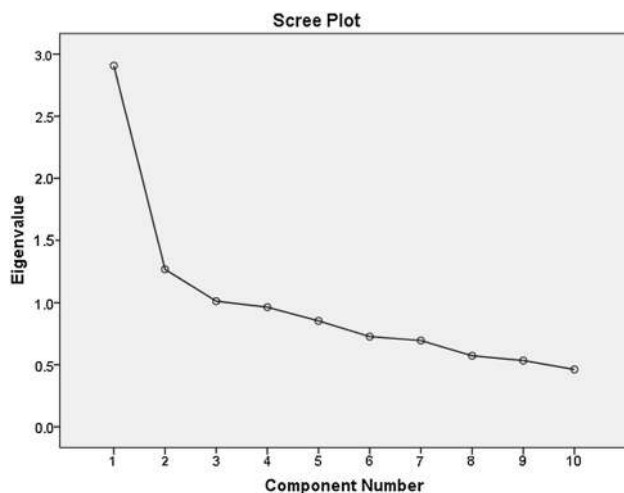
Fig. 1 Scree plot of principal component of POS items (N = 504)

Table 2 Rotated two-component matrix (orthogonal; N = 504)

POS items	Component		Conceptual domain of item
	1	2	
Anxiety	**0.772**		Psychological wellbeing
Depression	**0.658**	0.230	Psychological wellbeing
Family anxiety	**0.644**	−0.226	Psychological wellbeing
Pain	**0.585**	0.292	Physical
Symptoms	**0.575**	0.244	Physical
Feeling good	**0.567**	0.368	Psychological wellbeing
Time wasted	0.260		Quality of care
Information		**0.737**	Quality of care
Practical matters		**0.640**	Quality of care
Share feelings		**0.525**	Psychological wellbeing

Principal component analysis. Rotation: varimax with Kaiser normalization. Bold formatting indicates loadings ≥|0.400|

rotation (orthogonal vs. oblique), with all the items loading on the same components.

The results of PCA indicated that the POS consists of two domains that are moderately correlated. These domains do not appear to be consistent with predefined conceptual domains of the POS. Our findings suggest that the POS constitutes a measure with no clear multidimensionality. Thus, it was deemed necessary to conduct Rasch analysis on the whole instrument, rather than on any specific domain, in the next stage of the analysis.

3.2 Steps 2 and 3: Use of Rasch Analysis and Expert Opinion to Merge Categories, Eliminate Items and Develop a Unidimensional Scale

3.2.1 Item-Level Ordering

A total of nine items (items 1, 2, 4, 5, 6, 7, 8, 9 and 10) were disordered in the initial Rasch model. For two of the nine disordered items (item 1 'pain' and item 2 'other symptoms'), 'slightly' and 'moderately' were collapsed into a single category, as were 'severely' and 'overwhelmingly', resulting in three categories per item. Similarly, 'family anxiety', 'shared feelings', 'depression' and 'feeling good' (items 4, 6, 7 and 8, respectively) were converted to three-level items by merging 'occasionally' with 'sometimes' into a single category and 'most of the time' with 'always'. Wasted time (item 9) and practical matters (item 10), which have three levels in the original questionnaire, were converted to two-level items by merging 'half a day' with 'more than half a day' (item 9), and 'practical problems being addressed' with 'no practical problems' (item 10). The threshold probability curves for item 5 (information) suggested that this item would only work with two categories. Therefore, 'full information', 'information given but hard to understand', 'information given on request' and 'very little information given' were collapsed into a single category. However, because this merging was not deemed to be clinically meaningful, item 5 was eliminated from further analysis.

3.2.2 Rasch Model Goodness of Fit

After all thresholds were ordered, we assessed goodness of fit by examining overall and individual item statistics.

Table 3 Results of initial Rasch analysis of Palliative Care Outcome Scale (POS)-E (all items included)

Item		Threshold	Statistics after threshold re-ordering			
			Residual[a]	X-square	p value[b]	DIF
1	Pain	**Disordered**	–0.574	8.352	0.499	No
2	Other symptoms	**Disordered**	–1.410	21.811	0.010	No
3	Anxiety	Ordered	**–3.254**	40.843	**0.000**	No
4	Family anxiety	**Disordered**	–0.046	10.655	0.300	No
5	Information	**Disordered**	**3.442**	46.423	**0.000**	**Yes**
6	Shared feelings	**Disordered**	**3.758**	34.484	**0.000**	No
7	Depression	**Disordered**	–1.237	9.849	0.363	No
8	Feeling good	**Disordered**	–1.048	8.598	0.475	No
9	Time wasted	**Disordered**	2.177	25.787	0.002	**Yes**
10	Practical matters	**Disordered**	1.118	11.222	0.261	No
Overall model statistics after threshold re-ordering			Total item X-square = 218.025; p = 0.0000			
			Person-separation index: 0.657			

All statistics showing item misfit into the Rasch model are presented in bold

DIF differential-item functioning

[a] Residuals >2.5 or <−2.5 are considered high

[b] $p < 0.01$ indicates items that do not meet Rasch item fit criteria

Initial overall fit statistics of the items indicated poor fit to the Rasch model, with items 3, 5 and 6 showing misfit (a fit residual beyond ±2.5 and a chi-squared probability significant at the 0.001 level). Items 5 and 9 also exhibited DIF. Results of the initial analysis on all items are shown in Table 3. Based on the results of Rasch analysis, a number of items were consecutively excluded from further analysis according to our exclusion criteria until a good model fit was achieved.

Successive Rasch analyses led to the exclusion of items 5, 6 and 9 as they persistently had a poor fit to the Rasch model. For example, item 5 (information) had the poorest fit when compared with other items, it exhibited DIF, and its thresholds could only be ordered by combining adjacent levels in a way that was neither cognitively nor clinically meaningful. Items were excluded one at a time and both Rasch statistics and the PSI were constantly checked. This resulted in a final scale consisting of seven items (1, 2, 3, 4, 7, 8 and 10). With the exception of item 10, all other items had three response levels (e.g. 'not at all', 'occasionally or sometimes' and 'most of the time or always'). Item 10 (which originally had three levels) was collapsed to two levels: 'no problems or problems resolved' and 'problems in the process of being resolved or problems exist' (Table 4). The scale demonstrated a good model fit (X^2 probability 0.047). All items had a reasonable fit, as shown in Table 5, and no DIF was observed. The PSI reached a reasonable level of 0.678.

Table 4 Items and levels in final Palliative Care Outcome Scale (POS)-E scale

Item	Categories	Score
Family anxiety	No, not at all	0
	Occasionally/sometimes	1
	Most of the time/always	2
Other symptoms	No, not at all	0
	Slightly/moderately	1
	Severely/overwhelmingly	2
Pain	No, not at all	0
	Slightly/moderately	1
	Severely/overwhelmingly	2
Depression	No, not at all	0
	Occasionally/sometimes	1
	Most of the time/always	2
Anxiety	No, not at all	0
	Slightly/moderately	1
	Severely/overwhelming	2
Practical matters	Addressed	0
	Not addressed	1
Feeling good	Always/most of the time	0
	Occasionally/sometimes	1
	Not at all	2

Figure 2 shows the threshold map with items arranged in order of increasing difficulty from top to bottom, and with severity levels increasing from left to right.

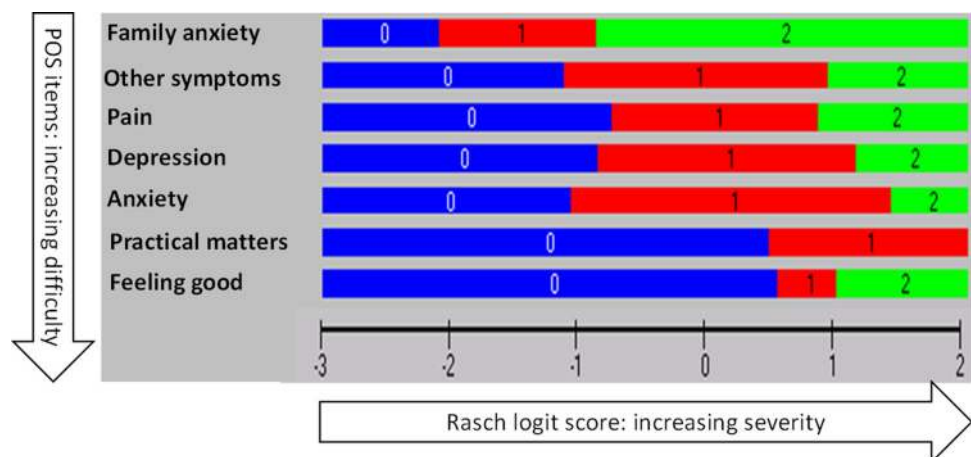**Table 5** Rasch statistics of the Palliative Care Outcome Scale (POS)-E measure

Item		Rasch analysis statistics		
		Residual	X-square	p value
1	Pain	0.452	5.586	0.694
2	Other symptoms	–0.424	11.073	0.198
3	Anxiety	–2.090	20.088	0.010
4	Family anxiety	1.221	11.423	0.179
7	Depression	0.247	10.893	0.208
8	Feeling good	1.084	6.422	0.600
10	Practical matters	2.951	9.339	0.315
Overall model statistics		Total item X-square = 74.825; p = 0.0472		
		Person-separation index: 0.678		

As shown in Fig. 3, the item map demonstrates that the new instrument is well targeted to the study population as it is able to capture the whole range of severity of palliative-care symptoms, with minimal floor or ceiling effects and good spread of items across the full range of respondents' scores.

3.2.3 Deriving Plausible Health States From the POS-E for Utility Measurement

The threshold map (Fig. 2) was used to derive plausible health states. This map illustrates the most likely combinations of item responses expected to be obtained by the study population at various levels (locations) of symptom severity. Items have been ordered from the easiest (item 4 'family anxiety') to the most difficult (item 8 'feeling good'), as indicated by their average location in the Rasch model. Shaded areas 0 (blue), 1 (red) and 2 (green) correspond to the three levels 'not at all', 'occasionally or sometimes' and 'most of the time or always', respectively,

with the exception of item 10, which has two levels: 0 (no problems or problems resolved) and 1 (problems in the process of being resolved or problems exist). The threshold map allows prediction of the most likely responses at various levels of severity. For example, a person whose symptom severity corresponds to location 0 on the logit scale is expected to most likely respond 0011112 (to items 8, 10, 3, 7, 1, 2, and 4, respectively).

Each combination of item responses represents a plausible health state likely to be observed in people with common palliative care problems. As illustrated in Table 6, a total of 14 distinct health states can be identified.

The results of the test for unidimensionality proposed by Smith [48] showed that the proportion of independent t tests that were significant at the 0.05 level was 1.52% (well below the 5% level), thus supporting the unidimensionality of the classification system.

3.3 Step 4: Validation of the Classification System

The POS-E was validated on the validation sample ($N = 508$): the scale had satisfactory overall and item fit statistics and no DIF was observed. The post hoc unidimensionality test also verified the scale's unidimensionality in this sample, and the threshold map indicated the same most likely item–response combinations (reflecting plausible health states) with those demonstrated by the analyses on the estimation sample. In total, the POS-E describes 1458 health states.

4 Discussion

We describe the first stage in developing a health-state classification for palliative care: the POS-E. Using rigorous research methods [8], we have derived the POS-E

Fig. 2 Threshold map illustrating plausible health states obtained by Rasch analysis. *POS* Palliative Care Outcome Scale

Rasch threshold map illustrating items in order of increasing difficulty from top to bottom, and severity levels increasing from left to right

Fig. 3 Item map of the
Palliative Care Outcome Scale
(POS)-E showing the
distribution of items across
respondents

Rasch item map showing the distribution of item-levels across respondents
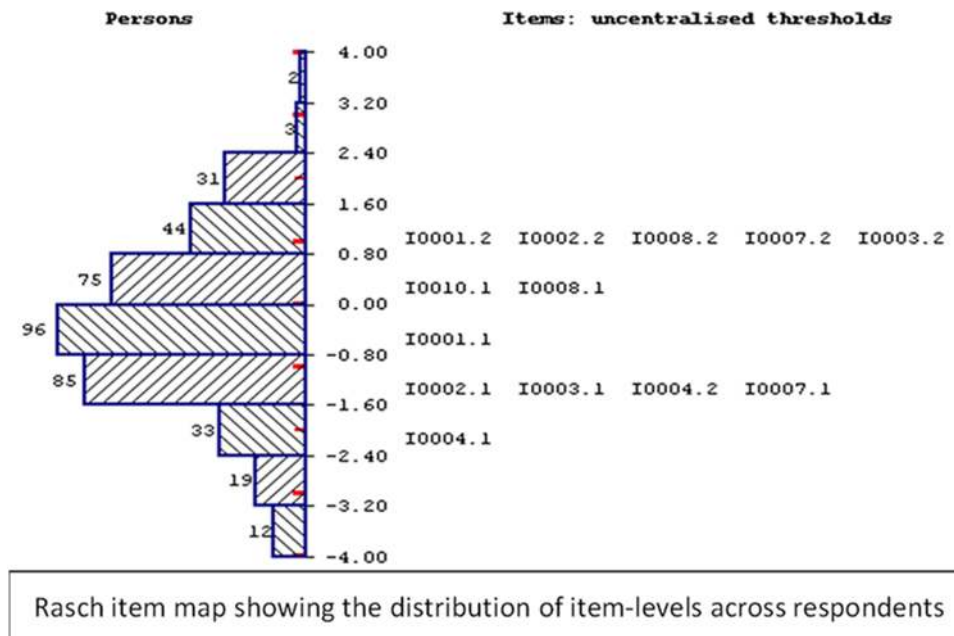
Table 6 Health states (and
coverage) of the Palliative Care
Outcome Scale (POS)-E as
identified by the threshold map

Item		Health states ($N = 504$)													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
4	Family anxiety	0	1	1	1	2	2	2	2	2	2	2	2	2	2
2	Other symptoms	0	0	1	1	1	1	1	1	1	1	2	2	2	2
1	Pain	0	0	0	0	0	0	1	1	1	2	2	2	2	2
7	Depression	0	0	0	0	0	1	1	1	1	1	1	1	2	2
3	Anxiety	0	0	0	1	1	1	1	1	1	1	1	1	1	2
10	Practical matters	0	0	0	0	0	0	0	1	1	1	1	1	1	1
8	Feeling good	0	0	0	0	0	0	0	0	1	1	1	2	2	2
Coverage (N)		40	79	13	49	17	8	128	35	14	13	27	14	21	46

classification system from an existing palliative care
measure, the POS. The next stage of the research will
involve preference elicitation and related regression-based
statistical modelling to derive preference weights for all
health states described by the POS-E. This will result in a
CSPBM capable of generating QALYs for use in economic
evaluations in palliative care.

POS-E is a unidimensional seven-item scale able to
capture the full range of severity of palliative care needs.
Six of the items have three levels each, and one item
(measuring practical matters) has two levels. The PSI of
this scale was approximately 0.68, which is somewhat
lower than the 0.70 value generally considered accept-
able for group comparison [50]. Nevertheless, 0.68 was
deemed adequate for our purpose, given the ability of the
scale to discriminate amongst different respondent groups
needed to be traded off with its conciseness and conve-
nience in a valuation survey, wherein respondents need to

process a combination of individual statements rather than
a summated scale score.

One limitation of our approach, similar to the method-
ology proposed by Sugar et al. [51], is that the number of
generated health states is limited and does not capture the
whole range of plausible combinations of responses.
Despite generating a limited number of health states,
application of this approach allows for the valuation of all
potential health states described by the POS-E. An
advantage of Rasch analysis over the clustering-based
approach is that it assigns all potential health states (i.e. all
combinations of item responses including those not illus-
trated in threshold maps) to different locations along the
scale according to their level of severity. The relationship
between the location of the health states across the latent
variable and the respective utility values obtained in a
valuation exercise can be estimated and used to generate
utility values for all patients completing POS-E. This

solution has been explored using regression techniques in a subsequent application of this approach on the Flushing questionnaire [52]. The findings of this latter study showed it is possible to assign appropriate utility values to all potential health states of a measure based on their location along the latent variable as estimated by Rasch analysis. However, it is conceivable that the Rasch approach we used would be best suited to a unidimensional instrument.

Developing a CSPBM from an existing palliative care measure has numerous advantages. Adapting a widely accepted and commonly used instrument such as the POS enables retrospective analysis of existing datasets and increases the likelihood that the measure will be used in future studies [24].

However, a major disadvantage of CSPBMs is that they may be prone to focusing effects where the effect of the condition is overrated because respondents to the valuation survey focus solely on the areas of health included in the classification system rather than viewing them in a broader perspective. Another disadvantage of CSPBMs is the correlation between perfect health and the best possible state described by a classification system. It is conceivable that a person could endorse the best possible health state based on a specific instrument but still have other problems not covered by its classification system. Thus, it becomes challenging to compare results between different PBMs because 'best possible' health states are instrument specific [8].

Nevertheless, these disadvantages are perhaps less crucial when the condition of interest is the overriding factor in determining HRQoL, as is likely to be the case for patients receiving palliative or end-of-life care. Furthermore, because advanced life-limiting conditions affect people's HRQoL in a wide variety of ways, the POS-E classification system covers a wider range of dimensions than many other CSPBMs. The decision on whether to use a CSPBM or a generic PBM will always involve a trade-off between the pros and cons of CSPBMs relative to the condition of interest [8]. In the case of palliative and end-of-life care, the potential limitation of existing generic measures [13], the wide range of the POS-E classification system, and the likely dominant nature of palliative care needs in determining HRQoL all favour the development and use of a CSPBM. The argument in favour of CSPBMs for palliative care is further strengthened by research around the role of capabilities and wellbeing in end-of-life care, which highlights that the objectives of end-of-life care do not always focus solely on health but may also include impacts on wellbeing [53]. This is particularly evident in the development work for the ICECAP Supportive Care Measure (ICECAP-SCM) [54], which is a CSPBM that measures capability at the end of life for use in economic evaluations. The POS-E relates to the ICECAP-SCM in that both instruments seek to incorporate important aspects of palliative and end-of-life care into economic evaluations. Standard economic instruments have been criticised for failing to do this [10, 11] However, there are important differences between the two instruments, mainly due to conceptual differences in their respective evaluative frameworks. The POS-E measures impact on health (or utility), whereas the ICECAP-SCM gives more attention to broader impacts on capability and wellbeing and is particularly important where health outcomes are not the focus of evaluation, such as social care interventions [55]. Nevertheless, because palliative and end-of-life care include aspects of both health (e.g. pain) and wellbeing (e.g. availability of social support), among other things, the POS-E and ICECAP-SCM can be regarded as complementary rather than mutually exclusive. Our analysis is based on pooled data from six studies, which was necessary to obtain a large enough sample to produce reliable and representative estimates. However, because the data were from patients with different types of cancer and those without cancer, it is perhaps a reasonable reflection of the diverse diagnoses of palliative care patients and therefore arguably more generalizable.

5 Conclusion

This study has shown that reducing the POS to a health-state classification system for palliative care (POS-E) is possible and that the results are robust. The POS-E classifies palliative care states as a combination of seven items: pain, other symptoms, anxiety, depression, family anxiety, feeling good about oneself, and practical matters. We also identified 14 plausible health states that can be used to value the HRQoL of patients receiving palliative care.

6 Further Research

The next step for this study is to undertake a valuation survey to attach appropriate utility values to all health states of the POS-E and thus convert it into a preference-based index. Our aim is that the new PBM will be suitable for cost-utility analyses of palliative care interventions where the use of generic PBMs such as the EQ-5D has been shown to be problematic [56–58]. Since this measure has been derived from the POS, an instrument routinely used for outcome monitoring in patients receiving palliative care in the UK and beyond, this study is expected to enable wider assessment of healthcare interventions for managing patients receiving palliative care in the form of cost-utility analysis.

Appendix

See Tables 7, 8 and 9.

Descriptive Statistics of Datasets

Dataset 1: a cancer mortality follow-back survey ($N = 596$) from 2009 to 2010 in London (The QUALYCARE study).

Dataset 2: a cross-sectional study on symptom burden and palliative care needs in chronic obstructive pulmonary disease and cancer in Germany ($N = 109$).

Dataset 3: a study of Parkinson's disease (longitudinal community study of predictive factors; $N = 82$).

Dataset 4: a randomised phase II trial of dignity therapy ($N = 45$, UK).

Dataset 5: a longitudinal study on trajectories of illness of stage 5 chronic renal disease ($N = 74$, UK).

Table 7 Descriptive statistics for development and validation datasets

Characteristic	Development ($N = 504$)	Validation ($N = 508$)
Female	53%	52%
Male	47%	48%
Age <60 years	16%	12%
Age >60 years	84%	88%
Cancer	79%	77%
Non-cancer	21%	23%
Mean total POS score	13.03	13.14

POS Palliative Care Outcome Scale

Table 8 Rotated three-component matrix[a] ($N = 504$)

Item	Component			Conceptual domain of item
	1	2	3	
Anxiety	**0.764**			Psychological wellbeing
Depression	**0.707**			Psychological wellbeing
Feeling good	**0.645**	0.207	–0.265	Psychological wellbeing
Family anxiety	**0.625**	–0.305		Psychological wellbeing
Pain	**0.600**	0.259		Physical
Symptoms	**0.543**	0.309	0.327	Physical
Information		**0.755**		Quality of care
Practical matters	0.211	**0.697**		Quality of care
Share feelings	0.240	0.283	**–0.667**	Psychological wellbeing
Time wasted			**0.602**	Quality of care

Extraction method: principal component analysis. Rotation method: varimax with Kaiser normalization. Bold formatting indicates loadings $\geq |0.400|$

[a] Rotation converged in five iterations

Table 9 Descriptive statistics of datasets according to study

Characteristic	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
N	596	109	82	45	74	105
Mean age ± SD	74 ± 12.8	65 ± 9.56	67 ± 8.82	67 ± 16.73	80 ± 6.74	67 ± 9.87
Female (%)	49	52	37	51	49	42
Ethnicity (%)						
White	92	95	80.4	84.4	68.9	77.1
Black	2.8	1	3.7	13.2	16.2	14.3
South Asian	1.7	2	4.8	2.4	8.1	1.7
Chinese	0.5	0	1.2	0	2.5	0
Other	2.6	2	6.1	0	4.3	5.7
Diagnosis (%)						
Cancer	100	45	0	100	0	20
Non-cancer	0	55	100	0	100	80
Mean POS total score ± SD	13.5 ± 6.72	10.3 ± 5.9	13.9 ± 6.2	11.7 ± 6.7	11.2 ± 7.1	15.0 ± 6.6

SD standard deviation

Dataset 6: an RCT on the effectiveness of an integrated palliative and respiratory care service for patients with advanced disease and refractory breathlessness in the UK in 2014 ($N = 105$).

References

1. Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? Health Econ. 2010;19(2):125–9. doi:10.1002/hec.1580.
2. Whitehead SJ. Health outcomes in economic evaluation: The QALY and utilities. Br Med Bull. 2010;96(1):5–21.
3. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095–108.
4. Hays RD, Revicki DA, Feeny D, Fayers P, Spritzer KL, Cella D. Using linear equating to map PROMIS global health items and the PROMIS-29 V2.0 profile measure to the health utilities index mark 3. Pharmacoeconomics. 2016;34(10):1015–22. doi:10.1007/s40273-016-0408-x.
5. Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med. 2005;60(7):1571–82. doi:10.1016/j.socscimed.2004.08.034.
6. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. Health Econ. 2002;11(5):447–56. doi:10.1002/hec.688.
7. 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. doi:10.1371/journal.pone.0151905.
8. 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 Technol Assess. 2012;16(32):1–114. doi:10.3310/hta16320.
9. World Health Organisation. Cancer pain relief and palliative care. Report of a WHO expert committee. WHO Technical Report Series. 1990; no. 804.
10. Normand C. Measuring outcomes in palliative care: limitations of QALYs and the road to PalYs. J Pain Symptom Manag. 2009;38(1):27–31.
11. Hughes J. Palliative care and the QALY problem. Health Care Anal. 2005;13(4):289–301.
12. Higginson IJ, Bausewein C, Reilly CC, Gao W, Gysels M, Dzingina M, et al. An integrated palliative and respiratory care service for patients with advanced disease and refractory breathlessness: a randomised controlled trial. Lancet Respir Med. 2014;2(12):979–87. doi:10.1016/S2213-2600(14)70226-7.
13. Round J. Is a QALY still a QALY at the end of life? J Health Econ. 2012;31(3):521–7.
14. Hearn J, Higginson IJ. Development and validation of a core outcome measure for palliative care: the palliative care outcome scale. Palliative Care Core Audit Project Advisory Group. Qual Health Care. 1999;8(4):219–27.
15. Bausewein C, Booth S, Gysels M, Kuhnbach R, Haberland B, Higginson IJ. Understanding breathlessness: cross-sectional comparison of symptom burden and palliative care needs in chronic obstructive pulmonary disease and cancer. J Palliat Med. 2010;13(9):1109–18. doi:10.1089/jpm.2010.0068.
16. Goodwin DM, Higginson IJ, Myers K, Douglas HR, Normand CE. Effectiveness of palliative day care in improving pain, symptom control, and quality of life. J Pain Symptom Manag. 2003;25(3):202–12.
17. Hall S, Petkova H, Tsouros AD, Costantini M, Higginson IJ. Palliative care for older people: better practices. Geneva: WHO; 2011. http://www.euro.who.int/en/publications/abstracts/palliative-care-for-older-people-better-practices. Accessed 20 June 2016.
18. Antunes B, Murtagh F, Bausewein C, Harding R, Higginson IJ. Screening for depression in advanced disease: psychometric properties, sensitivity, and specificity of two items of the Palliative Care Outcome Scale (POS). J Pain Symptom Manag. 2015;49(2):277–88. doi:10.1016/j.jpainsymman.2014.06.014.
19. Collins ES, Witt J, Bausewein C, Daveson BA, Higginson IJ, Murtagh FE. A systematic review of the use of the palliative care outcome scale and the support team assessment schedule in

palliative care. J Pain Symptom Manag. 2015;50(6):842–53.e19. doi:10.1016/j.jpainsymman.2015.07.015.

20. Costantini M, Rabitti E, Beccaro M, Fusco F, Peruselli C, La Ciura P, et al. Validity, reliability and responsiveness to change of the Italian palliative care outcome scale: a multicenter study of advanced cancer patients. BMC Palliat Care. 2016;15:23. doi:10.1186/s12904-016-0095-6.

21. Lowther K, Simms V, Selman L, Sherr L, Gwyther L, Kariuki H, et al. Treatment outcomes in palliative care: the TOPCare study. A mixed methods phase III randomised controlled trial to assess the effectiveness of a nurse-led palliative care intervention for HIV positive patients on antiretroviral therapy. BMC Infect Dis. 2012;12:288. doi:10.1186/1471-2334-12-288.

22. van Vliet LM, Harding R, Bausewein C, Payne S, Higginson IJ. How should we manage information needs, family anxiety, depression, and breathlessness for those affected by advanced disease: development of a Clinical Decision Support Tool using a Delphi design. BMC Med. 2015;13:263. doi:10.1186/s12916-015-0449-6.

23. Smith S, Brick A, O'Hara S, Normand C. Evidence on the cost and cost-effectiveness of palliative care: a literature review. Palliat Med. 2014;28(2):130–50. doi:10.1177/0269216313493466.

24. Brazier JE, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. UK: Oxford University Press; 2007.

25. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21(2):271–92.

26. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care. 2004;42(9):851–9.

27. Brazier JE, Roberts J, Platts M, Zoellner YF. Estimating a preference-based index for a menopause specific health quality of life questionnaire. Health Qual Life Outcomes. 2005;3:13. doi:10.1186/1477-7525-3-13.

28. Stevens KJ, Brazier JE, McKenna SP, Doward LC, Cork MJ. The development of a preference-based measure of health in children with atopic dermatitis. Br J Dermatol. 2005;153(2):372–7. doi:10.1111/j.1365-2133.2005.06736.x.

29. Brazier J, Czoski-Murray C, Roberts J, Brown M, Symonds T, Kelleher C. Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. Med Decis Making. 2008;28(1):113–26. doi:10.1177/0272989X07301820.

30. 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. doi:10.1177/0272989x12464431.

31. Rowen D, Brazier J, Young T, Gaugris S, King MT, Craig B, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Value Health. 2011;14(5):721–31. doi:10.1016/j.jval.2011.01.004.

32. Gomes B, McCrone P, Hall S, Koffman J, Higginson I. Variations in the quality and costs of end-of-life care, preferences and palliative outcomes for cancer patients by place of death: the QUALYCARE study. BMC Cancer. 2010;10(1):400.

33. Higginson IJ, Gao W, Saleem TZ, Chaudhuri KR, Burman R, McCrone P, et al. Symptoms and quality of life in late stage parkinson syndromes: a longitudinal community study of predictive factors. PLoS One. 2012;7(11):e46327. doi:10.1371/journal.pone.0046327.

34. Murtagh FEM, Sheerin NS, Addington-Hall J, Higginson IJ. Trajectories of illness in stage 5 chronic kidney disease: a longitudinal study of patient symptoms and concerns in the last year

of life. Clin J Am Soc Nephrol. 2011;6(7):1580–90. doi:10.2215/cjn.09021010.

35. Hall S, Goddard C, Opio D, Speck PW, Martin P, Higginson IJ. A novel approach to enhancing hope in patients with advanced cancer: a randomised phase II trial of dignity therapy. BMJ Supp Palliat Care. 2011;1(3):315–21. doi:10.1136/bmjspcare-2011-000054.

36. Smith A, Rush R, Fallowfield L, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. BMC Med Res Methodol. 2008;8(1):33.

37. Albers G, Echteld MA, de Vet HC, Onwuteaka-Philipsen BD, van der Linden MH, Deliens L. Evaluation of quality-of-life measures for use in palliative care: a systematic review. Palliat Med. 2010;24(1):17–37. doi:10.1177/0269216309346593.

38. Collins ES, Witt J, Bausewein C, Daveson BA, Higginson IJ, Murtagh FEM. A systematic review of the use of the palliative care outcome scale and the support team assessment schedule in palliative care. J Pain Symptom Manag. 2015;50(6):842–53.e19. doi:10.1016/j.jpainsymman.2015.07.015.

39. Bausewein C, Le Grice C, Simon S, Higginson I, PRISMA obo. The use of two common palliative outcome measures in clinical care and research: a systematic review of POS and STAS. Palliat Med. 2011;25(4):304–13. doi:10.1177/0269216310395984.

40. Tennant A, Pallant JF. Unidimensionality matters! (A Tale of Two Smiths?). In: Rasch measurement transactions. 2006. http://www.rasch.org/rmt/rmt201c.htm. Accessed 16 May 2014.

41. Cerny BA, Kaiser HF. A study of a measure of sampling adequacy for factor-analytic correlation matrices. Multivariate Behav Res. 1977;12(1):43–7. doi:10.1207/s15327906mbr1201_3.

42. Bartlett MS. Tests of significance in factor analysis. Br J Stat Psychol. 1950;3(2):77–85. doi:10.1111/j.2044-8317.1950.tb00285.x.

43. Horn J. A rationale and test for the number of factors in factor analysis. Psychometrika. 1965;30(2):179–85. doi:10.1007/bf02289447.

44. Watkins M. Monte Carlo PCA for parallel analysis. 2002. http://edpsychassociates.com/Watkins3.html. Accessed 9 June 2016.

45. Kieffer K, editor. Orthogonal versus oblique factor rotation: a review of the literature regarding the pros and cons. Paper presented at the 27th Annual Meeting of the Mid-South Educational Research Association. New Orleans, LA; November 4–6, 1998.

46. 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. doi:10.1007/s11136-008-9428-0.

47. 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 Care Res. 2007;57(8):1358–62. doi:10.1002/art.23108.

48. Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas. 2002;3(2):205–31.

49. 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. doi:10.1002/art.23108.

50. Fisher WJ. Reliability, separation, strata statistics. Rasch Meas Trans. 1992;6(3):238.

51. Sugar CA, Sturm R, Lee TT, Sherbourne CD, Olshen RA, Wells KB, et al. Empirically defined health states for depression from the SF-12. Health Serv Res. 1998;33(4 Pt 1):911–28.

52. Young TA, Rowen D, Norquist J, Brazier JE. Developing preference-based health measures: using Rasch analysis to generate health state values. Qual Life Res. 2010;19(6):907–17. doi:10.1007/s11136-010-9646-0.

53. Goranitis I, Coast J, Day E, Copello A, Freemantle N, Seddon J, et al. Measuring health and broader well-being benefits in the context of opiate dependence: the psychometric performance of the ICECAP-A and the EQ-5D-5L. Value Health. 2016;19(6):820–8. doi:10.1016/j.jval.2016.04.010.

54. Sutton EJ, Coast J. Development of a supportive care measure for economic evaluation of end-of-life care using qualitative methods. Palliat Med. 2014;28(2):151–7. doi:10.1177/0269216313489368.

55. Round J. Care at the end of life: an economic perspective. 1st ed. Cham: Springer International Publishing; 2016.

56. Brazier J. Measuring and valuing mental health for use in economic evaluation. J Health Serv Res Policy. 2008;13(Suppl 3):70–5. doi:10.1258/jhsrp.2008.008015.

57. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res. 2008;17(3):485–95. doi:10.1007/s11136-008-9317-6.

58. van de Willige G, Wiersma D, Nienhuis FJ, Jenner JA. Changes in quality of life in chronic psychiatric patients: a comparison between EuroQol (EQ-5D) and WHOQoL. Qual Life Res. 2005;14(2):441–51.