






# A brief, patient- and proxy-reported outcome measure in advanced illness: Validity, reliability and responsiveness of the Integrated Palliative care Outcome Scale (IPOS)

*Palliative Medicine*  
2019, Vol. 33(8) 1045–1057  
© The Author(s) 2019



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0269216319854264  
journals.sagepub.com/home/pmj



Fliss EM Murtagh<sup>1,2</sup>, Christina Ramsenthaler<sup>2,3</sup>, Alice Firth<sup>2</sup>,  
Esther I Groeneveld<sup>2</sup>, Natasha Lovell<sup>2</sup>, Steffen T Simon<sup>4</sup>,  
Johannes Denzel<sup>3</sup>, Ping Guo<sup>2</sup>, Florian Bernhardt<sup>3</sup>, Eva Schildmann<sup>3</sup>,  
Birgitt van Oorschot<sup>5</sup>, Farina Hodiamont<sup>3</sup>, Sabine Streitwieser<sup>3</sup>,  
Irene J Higginson<sup>2</sup> and Claudia Bausewein<sup>3</sup>

## Abstract

**Background:** Few measures capture the complex symptoms and concerns of those receiving palliative care.

**Aim:** To validate the Integrated Palliative care Outcome Scale, a measure underpinned by extensive psychometric development, by evaluating its validity, reliability and responsiveness to change.

**Design:** Concurrent, cross-cultural validation study of the Integrated Palliative care Outcome Scale – both (1) patient self-report and (2) staff proxy-report versions. We tested construct validity (factor analysis, known-group comparisons, and correlational analysis), reliability (internal consistency, agreement, and test–retest reliability), and responsiveness (through longitudinal evaluation of change).

**Setting/participants:** In all, 376 adults receiving palliative care, and 161 clinicians, from a range of settings in the United Kingdom and Germany

**Results:** We confirm a three-factor structure (Physical Symptoms, Emotional Symptoms and Communication/Practical Issues). Integrated Palliative care Outcome Scale shows strong ability to distinguish between clinically relevant groups; total Integrated Palliative care Outcome Scale and Integrated Palliative care Outcome Scale subscale scores were higher – reflecting more problems – in those patients with ‘unstable’ or ‘deteriorating’ versus ‘stable’ Phase of Illness ( $F = 15.1$ ,  $p < 0.001$ ). Good convergent and discriminant validity to hypothesised items and subscales of the Edmonton Symptom Assessment System and Functional Assessment of Cancer Therapy–General is demonstrated. The Integrated Palliative care Outcome Scale shows good internal consistency ( $\alpha = 0.77$ ) and acceptable to good test–retest reliability (60% of items  $k_w > 0.60$ ). Longitudinal validity in form of responsiveness to change is good.

**Conclusion:** The Integrated Palliative care Outcome Scale is a valid and reliable outcome measure, both in patient self-report and staff proxy-report versions. It can assess and monitor symptoms and concerns in advanced illness, determine the impact of healthcare interventions, and demonstrate quality of care. This represents a major step forward internationally for palliative care outcome measurement.

## Keywords

Outcome and process assessment, validation studies, psychometrics, reliability, patient-reported outcome measures, palliative care, symptom assessment

<sup>1</sup>Wolfson Palliative Care Research Centre, Hull York Medical School, University of Hull, Hull, UK

<sup>2</sup>Cicely Saunders Institute of Palliative Care, Policy & Rehabilitation, King’s College London, London, UK

<sup>3</sup>Department of Palliative Medicine, University Hospital, Ludwig-Maximilians-University Munich, Munich, Germany

<sup>4</sup>Center for Palliative Medicine, University of Cologne, Cologne, Germany

<sup>5</sup>Interdisciplinary Centre for Palliative Medicine, University Hospital Wuerzburg, Wuerzburg, Germany

## Corresponding author:

Fliss EM Murtagh, Wolfson Palliative Care Research Centre, Hull York Medical School, University of Hull, Allam Medical Building, Hull HU6 7RX, UK.

Email: fliss.murtagh@hyms.ac.uk

**What is already known about the topic?**

- Validated measures of symptoms and quality-of-life already exist for palliative care.
- Symptom measures do not capture broader concerns of those with advanced illness, such as information needs, family, practical issues, and so on.
- Quality-of-life measures are often weighted towards symptoms and function, and may be less relevant and applicable in far-advanced illness.
- Few existing measures have a validated proxy-report version.

**What this paper adds?**

- The Integrated Palliative care Outcome Scale (IPOS) is shown to be clinically meaningful, with good validity, reliability, and responsiveness to change.
- Both patient self-report and staff proxy-report versions of IPOS are evaluated, allowing maximum flexibility for clinical use.

**Implications for practice**

- IPOS is a comprehensive and psychometrically robust measure underpinned by what patients report as their main concerns, and brief enough to be used in advanced illness.
- This is a major advance for palliative care outcome measurement internationally.

**Background**

Healthcare systems around the world face major challenges because of increasing numbers of older people with multi-morbidities, and growing need for palliative care.<sup>1</sup> The increase in chronic diseases – now accounting for a third of all deaths globally – as well as population ageing, is responsible for these changes.<sup>2,3</sup> More people need palliative care,<sup>3</sup> with some estimates exceeding 40 million/year worldwide.<sup>1</sup>

To meet current challenges, unexplained variations in healthcare quality need to be addressed through improving outcomes.<sup>4</sup> This can only be achieved with measurement of individual patient-centred outcomes.<sup>5</sup> Patient-reported outcome measures (PROMs) are validated questionnaires completed by patients to measure their perceptions of their own health status/wellbeing.<sup>6</sup> Available PROMs tend to be illness-specific,<sup>7–10</sup> rather than capturing common concerns in advanced illness, regardless of diagnosis. Most PROMs focus on disease control or complications,<sup>11</sup> rather than the concerns of patients.

One of the few outcome measures which does capture the full range of concerns prioritised by those with advanced illness themselves, is the Palliative care Outcome Scale (POS).<sup>12</sup> POS was developed >15 years ago, has been psychometrically well tested,<sup>13–16</sup> and is widely used.<sup>17–19</sup> It is a brief, person-centred outcome measure,<sup>20</sup> yet incorporates the main concerns that people with advanced illness themselves prioritise.<sup>21</sup>

Although evidence from POS users demonstrated its value in practice and research, there was need for a more refined version (e.g. incorporating more symptoms, refining the spiritual/existential item for diverse populations)

to aid utility.<sup>22,23</sup> Having cognitively tested a refined version of POS, the Integrated Palliative care Outcome Scale (IPOS),<sup>24</sup> we therefore aimed to evaluate the validity, reliability, and responsiveness to change of the patient self-report and staff proxy-report IPOS. IPOS covers patients' main concerns, common symptoms, patient/family distress, existential well-being, sharing feelings with family or friends, information received, and practical concerns, within a timeframe of 3 days (for inpatient settings) or 7 days (for ambulatory settings).

**Methods**

We are following established quality criteria<sup>25</sup> to refine and validate IPOS. The first steps, involving cognitive testing, are published;<sup>24</sup> this involved cultural adaptation and cognitive interviewing in both English and German,<sup>24</sup> to establish a final IPOS in both languages (available for download at [www.pos-pal.org](http://www.pos-pal.org)). Here, we report testing of the construct validity, reliability, and responsiveness of this final version of IPOS, in both languages. The design is a multi-centre validation study of two versions of IPOS – (a) patient self-report and (b) staff proxy-report.

**Population and settings**

Patients receiving palliative care were consecutively recruited from eight UK and five German sites: (a) three UK and one German hospital consultation services, (b) five UK and three German in-patient palliative units, and (c) seven UK and two German community (home-based) palliative services. Staff caring for participating patients were also recruited. Data were gathered between June 2014 and January 2016.

*Inclusion/exclusion criteria:* Inclusion criteria for patient participants were: >18 years of age, capacity for written informed consent (as judged by the clinical team), and able to speak/read English or German. Exclusion criteria for patient participants were as follows: impaired capacity, too unwell or distressed to participate (as judged by their clinical teams); or unable to understand English or German. Inclusion criterion for staff participants was delivering care for a patient participant. Staff participants scored the research measures independently of the corresponding patient participant.

### Data collection

Demographic/clinical information for patient participants at baseline included age, gender, ethnicity, marital status, if living alone, presence/absence of family caregiver(s), performance status using Australia-modified Karnofsky Performance Status (AKPS),<sup>26</sup> and primary diagnosis.

Data (see Table 1 for data collection measures) were collected at two time-points; 2–5 days apart within inpatient settings, and 7–21 days apart in community settings. For patient participants, IPOS (patient version, 3-day recall period), Edmonton Symptom Assessment System–revised (ESASr)<sup>30–32</sup> and the Functional Assessment of Cancer Therapy-General (FACT-G)<sup>33</sup> were collected at first time-point, and IPOS (patient-version) and global change question (see Table 1)<sup>34,40</sup> were collected at the second time-point. For staff participants, the staff-version of IPOS, the Support Team Assessment Schedule (STAS),<sup>41</sup> AKPS<sup>26</sup> and Phase of Illness<sup>38</sup> were collected at the first time point, and IPOS (staff-version), AKPS, and Phase of Illness were collected at the second time point.

### Analysis

All data were independently double-entered and cross-checked. If not otherwise stated, missing data were excluded pairwise for all analyses.

*Descriptive statistics* were used to describe the sample, and the range/distribution of scores for individual IPOS items, IPOS subscales (as derived from the factor analysis), and total scores.

### Construct validity

(a) *Structural validity:* We undertook confirmatory factor analysis (CFA) to establish structural validity and subscales. Taking prior factor analyses into account,<sup>15,16</sup> we contrasted a 2-factor with a 3-factor solution. We used robust maximum likelihood estimation to accommodate the ordinal nature of the data.<sup>42</sup> Fit of each solution was evaluated using chi-square, ratio of chi-square and degrees of freedom, confirmatory fit index, Tucker-Lewis index

and root mean square error of approximation (RMSEA).<sup>43</sup> Contrasting models were compared regarding fit indices, standardized parameter estimates, and local strains (low loadings, high standard error).<sup>44</sup>

(b) *Known-group comparisons:* We hypothesized from prior evidence<sup>38</sup> that

- Those with ‘unstable’ or ‘deteriorating’ Phase of Illness would have (i) a higher total IPOS score (more symptoms/concerns) and (ii) higher physical symptom scores on IPOS, than those with ‘stable’ Phase of Illness.
- Those with lower function (AKPS) would have (iii) higher total IPOS scores and (iv) higher physical symptom scores on IPOS.<sup>45</sup>

Non-parametric tests after checking of assumptions were used, Kruskal–Wallis H test for hypotheses (i) and (ii), and Mann-Whitney U test for (iii) and (iv).

(c) *Convergent and discriminant validity* was tested by correlating individual IPOS items and subscales with respective items and subscales from patient-reported ESAS-r<sup>32</sup> and FACT-G<sup>33</sup> measures, using Spearman’s correlation coefficients with associated p-values. We hypothesized:

- High correlations ( $r > 0.70$ ) of identical or near-identical single items relating to the physical/psychological symptoms from ESAS and IPOS.
- Mid-range correlations ( $r \ 0.5 \leq 0.7$ ) between (i) total ESAS scores (which includes only symptoms) and (ii) FACT-G total/subscale scores (not covering spiritual, practical and family issues as in IPOS), with total IPOS scores (including domains beyond symptoms).

### Reliability

(a) *Internal consistency* was estimated using Cronbach’s  $\alpha$  for IPOS total scores and subscales. The normally accepted threshold for good internal consistency of 0.8<sup>25,46</sup> was lowered to 0.6 due to the multi-dimensional, non-redundant nature of the IPOS.<sup>47</sup>

(b) *Test–retest reliability* was determined among those patients reporting no change on the global change rating.<sup>34</sup>

(c) *Inter-rater reliability* was assessed between independent patient and staff ratings, and between two independent staff ratings. Cohen’s weighted kappa ( $\kappa_w$ ) was calculated as the reliability statistic, together with proportion agreement of cases where staff or patient’s ratings were equal to or within +1 or –1 of the score, and Spearman correlation to test the association between patient/

Table 1. Study measures.

Measure	Details of measure	Background, validation and references
The Integrated Palliative care (or Patient) Outcome Scale (IPOS) <i>In this study, reported by both patient and staff participants independently, at T1 and T2</i>	IPOS combines the items from the Palliative Care Outcome Scale (POS) and those from its symptom list (POS and POS-S) into one integrated measure. There are two versions of IPOS: patient self-report and staff proxy-report. The IPOS versions consist of 20 (patient version) or 19 (staff version) items: one free-text question about main problems and concerns, 17 items on physical, psychological, spiritual and problems, communication needs including with family, and practical support. They are scores on a 5-point Likert-type scale from 0 (best) to 4 (worst). One additional free-text item asks about additional symptoms to be specified and scored. In the patient version, there is also an additional item reporting by whom the measure was completed (patient alone, with family help, with staff help). Only the 17 standardised items contribute to the IPOS total score. The full IPOS measure is available for download at <a href="http://www.pos-pal.org">www.pos-pal.org</a> .	The original Palliative care Outcome Scale (POS) included 10 items covering domains most important to patients with advanced illness. <sup>12</sup> Following patient and staff feedback, a symptom module (POS-S), adapted for specific conditions) was added. Staff versions of POS and POS-S were also developed. <sup>27</sup> Both patient- and staff-reported versions of POS and POS-S have undergone extensive psychometric testing. Both measures show good validity and internal consistency <sup>12,20</sup> as well as test-retest reliability <sup>12</sup> in diverse settings. <sup>13–20</sup> Factor analysis suggests two subscales, psychological well-being and quality of care. <sup>15</sup> It has been translated, culturally adapted and validated for use in fourteen languages. <sup>16,17,28,29</sup> and is widely used internationally. <sup>18,19,23</sup> Phase 1 of this IPOS validation study included cognitive interviewing to assess acceptability and content/face validity, and identify cognitive processing issues; this has been reported previously. <sup>24</sup>
The Edmonton Symptom Assessment System revised (ESASr) <i>In this study, reported by patient participants, at T1</i>	ESAS consists of nine visual analogue scales, scored from 0–10, including pain, shortness of breath, nausea, depression, activity, anxiety, well-being, drowsiness and appetite. Initially, ESAS was developed to measure the most common symptoms in cancer patients. Higher scores indicate worse symptoms.	The revised ESAS has been widely validated for use in assessing the symptoms of patients with advanced progressive illness. <sup>30–32</sup>
The Function Assessment of Cancer Therapy – General (FACT-G) <i>In this study, reported by patient participants, at T1</i>	FACT-G comprises 27 items, divided into four primary quality of life domains: Physical well-being (7 items), Social/family well-being (7 items), Emotional well-being (6 items), and Functional well-being (7 items). Each item is measured on a 5-point Likert-type scale. Higher scores indicate better functioning.	Initially developed as a quality-of-life measure for evaluating patients receiving cancer treatment, it has been widely validated across long-term conditions. <sup>33</sup>
Global change question <i>In this study, reported by patient participants, at T2</i>	Single-item asking patient participants to report overall change in their symptoms and concerns; 'Over the last three days would you say that things have got better/worse /there has been no change'.	A single global 'change' question recommended for assessing responsiveness of patient-reported outcome measures. <sup>34</sup>
The Support Team Assessment Schedule (STAS) <i>In this study, reported by staff participants, at T1</i>	STAS represents the first available staff-rated palliative care clinical assessment, comprising 9 core and up to 20 optional items covering physical, psychosocial, spiritual, communication, planning, family concerns and service aspects. We used the supplementary definitions and ratings for individual symptoms as described in Clinical Audit in Palliative Care. <sup>35</sup>	It is a validated tool designed to allow clinical staff to assess the clinical and intermediate outcomes of palliative care. <sup>36,37</sup>
Phase of illness. <i>In this study, reported by staff participants, at T1 and T2</i>	Single item staff-reported measure to provide the context of the current phase of illness with four categories; stable, unstable, deteriorating, and terminal. This single item measure is recorded by staff.	Phase of illness categorises seriously ill patients according to the acuteness and urgency of palliative needs, and has been used as a predictor of resource use for Australian sub-acute and non-acute healthcare. <sup>38</sup> It shows good inter-rater reliability and clinical utility in populations with advanced progressive illness. <sup>38,39</sup>
Australia-modified Karnofsky Performance Status <i>In this study, reported by staff participants, at T1 and T2</i>	A single score between 0% and 100% (in 10% steps) based on a patient's ability to perform common tasks relating to activity, work and self-care. A score of 100% signifies normal physical abilities with no evidence of disease. Decreasing numbers indicate reduced performance status.	The Australia-modified Karnofsky Performance Status (AKPS) is based on the Karnofsky Performance Status, but is adapted for advanced illness. <sup>26</sup>

staff or two independent staff ratings. For interpretation, the Landis and Koch<sup>48</sup> and Fleiss<sup>49</sup> criteria of  $k > 0.4$  for fair to good and  $k > 0.75$  for substantial to excellent agreement were used.

**Responsiveness to change.** We assessed responsiveness using a distribution-based approach.<sup>50</sup> We compared mean changes and respective standard deviations of change descriptively in the six categories of change given by the global change rating (ranging from much better to much worse with a 'don't know'-category).

All analyses were conducted using SPSS version 24.0.<sup>51</sup> The R lavaan package (version),<sup>52</sup> was used for CFA. A  $p$ -value  $< 0.05$  was considered statistically significant for all analyses.

**Sample size.** Sample size considerations were based on guidelines for sample sizes for factor analysis<sup>53</sup> and a Monte Carlo study, determining the sample size to detect factor loadings of at least 0.40 (based on former factor analytic evidence from the POS),<sup>15,16</sup> a power of 80% and an alpha level of 0.05. Simulations were run using the R simsem package,<sup>54</sup> using 10,000 replications and assuming missing data to be handled within a full information maximum likelihood approach. A sample size of 320 was deemed sufficient for modelling.

## Results

### Subject characteristics

In all, 392 patient participants were recruited. Screened, eligible, approached and consented participants, and first ( $n = 376$ ) and second ( $n = 275$ ) timepoint completion, with reasons for non-completion, are shown in Appendix Figure 1. Demographic and clinical characteristics are reported in Table 2.

### Descriptive statistics and distribution

Table 3 shows prevalence for IPOS items, distribution of IPOS scores, and % of missing data, at first timepoint. The full range of response options was used; only the items 'Vomiting', 'Having enough information' and 'Practical matters' showed positive skew above  $\pm 1.0$ . There was little missing data; the highest percentage of missing data was for 'Poor appetite' (3.5%), 'Family anxiety' (2.4%), 'Vomiting' (2.1%), 'Practical matters' (2.1%), 'Having enough information' (1.9%), and 'Feeling at peace' (1.9%).

### Structural validity, identification of subscales, and internal consistency

An initial CFA was conducted to test for uni-dimensionality, a model using all 17 scorable IPOS items loading onto one

**Table 2.** Demographic and clinical characteristics for all patient participants ( $n = 376$ ).

Variable	Patients	
	n	%
<b>Setting</b>		
Hospital inpatient	180	47.9
Hospice inpatient	72	19.1
Hospital outpatient	5	1.3
Community (home-based)	95	25.3
Respite (in-patient)	13	3.5
Missing	11	2.9
<b>Country</b>		
Germany	154	40.4
United Kingdom	222	59.6
<b>Socio-demographic details</b>		
<b>Age</b>	Mean 65.8 (median: 67)	(SD 13.2; range 20–93)
<65 years	157	41.6
≥65 years	219	58.4
<b>Gender</b>		
Men	174	46.3
Women	187	49.7
Missing	15	4.0
<b>Ethnic origin</b>		
White	342	91
Black African or Black Caribbean	8	2.1
Asian	4	1.1
Mixed ethnic background	1	0.3
Missing	21	5.6
<b>Marital status</b>		
Single	36	9.6
Married	208	55.3
Divorced or separated	56	14.9
Widowed	57	15.2
Missing	19	5.1
Having a carer	202	53.7
Living alone	132	35.1
<b>Disease factors</b>		
<b>Phase of illness</b>		
Stable	164	43.6
Unstable	129	34.3
Deteriorating	52	13.8
Dying	1	0.3
Missing	30	8.0
<b>Primary diagnosis</b>		
Cancer	292	77.7
Digestive organs	82	21.8
Respiratory tract	47	12.5
Genitourinary tract	63	16.8
Breast	29	7.7
Lymph/	15	4.0
Haematopoietic		
Other cancers <sup>a</sup>	56	14.9
Non-cancer	57	15.2

(Continued)

**Table 2.** (Continued)

Variable	Patients	
	n	%
COPD	24	6.4
Stroke, MND	11	2.9
HIV/AIDS	2	0.5
Renal failure	3	0.8
Liver failure	7	1.9
Heart failure	2	0.5
Other <sup>b</sup>	8	2.2
Missing	27	6.7
Functional status		
<i>Australia-modified Karnofsky performance status</i>		
Mean (SD, range)	56.8	(SD 15.8; range 0–90)
0–50	150	39.8
60–100	219	58.2
Missing	7	1.8
IPOS completion		
Completed IPOS alone	162	43.1
Completed IPOS with family help	37	9.8
Completed IPOS with staff help	168	44.7
Missing	9	2.4
<i>Time between assessment 1 and 2 (in days)</i>	Mean 6.6 (median 4)	(SD 7.6; range 1–62)

COPD: chronic obstructive pulmonary disease; MND: motor neurone disease; IPOS: Integrated Palliative care Outcome Scale.

aOther cancers comprised cancers of lip/oral cavity/pharynx, skin, brain and central nervous system (CNS), and multiple sites.

bNon-specified or other non-cancer disease.

cNot staff participants in the study.

latent variable. As expected for IPOS (a multi-dimensional measure), the goodness-of-fit indices of this initial CFA suggest no adequate fit, with fit indices comparative fit index (CFI) (0.55) and RMSEA (0.11) well below and above the recommended cut-offs (see Appendix Table 1). The three-factor solution showed a better fit than the two-factor solution (see standardised factor loadings in Appendix Table 2). The first factor, Physical Symptoms, comprises 10 items and explains 24.9% of variance. The second factor, Emotional Symptoms, consists of 4 items and explains 12.3% of variance. The third factor, Communication/Practical Issues, contains 3 items, and explains 8.3% of variance. (These three factors we use throughout the analysis as subscales; the Physical, Emotional, and Communication/Practical subscales.) Total and subscale statistics are shown in Table 4. Cronbach's  $\alpha$  was 0.77, showing good internal consistency for the total scale. For the IPOS subscales, Cronbach's  $\alpha$  was 0.70, 0.68, and 0.58, respectively.

### Construct validity

One-way analysis of variance supported our prior hypothesis that total IPOS scores and IPOS subscale scores were higher in those patients with unstable or deteriorating Phase of Illness compared to stable Phase of Illness ( $F = 15.1$ ,  $p < 0.001$  for total IPOS and  $F = 17.8$  and  $5.7$ ,  $p < 0.003$  for IPOS Physical and IPOS Emotional symptoms, respectively) (see Appendix Figure 2).

The total IPOS ( $t = 2.8$ ,  $p = 0.006$ ), IPOS Physical Symptoms subscale ( $t = 3.8$ ,  $p < 0.001$ ), and individual IPOS items 'Shortness of breath', 'Weakness or lack of energy', 'Drowsiness', 'Poor mobility', 'Family anxiety', 'Depression' and 'Information' were all able to distinguish between those patients with higher versus lower functional status on Australia-modified Karnofsky Performance Status (60%–100% vs 0%–50%) (see Appendix Table 5). Because of skewed data, these comparisons were also run using equivalent non-parametric tests, with highly similar results.

Convergent and discriminant validity assessment also comprised testing a series of hypotheses for how IPOS subscales and single items correlate with single items, subscales and total scores of ESAS and FACT-G. Correlations were confirmed, being in the hypothesised range of magnitude and direction (see Appendix Tables 3 and 4).

### Reliability

In all, 66 patients self-classified as stable between the two timepoints. This was confirmed using the staff-reported 'Phase of Illness'. For these 66 stable patients, test-retest reliability weighted kappa values showed good to very good agreement (range 0.50–0.8), except for the items 'Sharing feelings with family or friends' ( $\kappa = 0.20$ ), 'Having enough information' ( $\kappa = 0.39$ ), 'Feeling at peace' ( $\kappa = 0.43$ ), and 'Drowsiness' ( $\kappa = 0.43$ ). The proportion agreement within one score between assessments was generally good to excellent with these four items being the only ones with proportions below 80% (see Table 5). Note that Cohen's weighted kappa was calculated using *all* answer options, and for each item of the IPOS, plus all subscale and total scores of the IPOS.

For the assessment of inter-rater agreement between two independent staff, a maximum of 95 matched pairs per IPOS item was available. Agreement as measured by weighted Kappa scores was good ( $\geq \kappa = 0.4$ ) for 11 of 17 IPOS items with the highest levels of agreement being achieved for the items 'Pain' ( $\kappa = 0.72$ ), 'Shortness of breath' ( $\kappa = 0.82$ ) and 'Nausea' ( $\kappa = 0.63$ ). Lower levels of agreement were observed for items 'Weakness or lack of energy', 'Drowsiness', 'Family anxiety', 'Sharing feelings with family or friends', 'Has the patient had enough information as s/he wanted?' and 'Have any practical matters resulting from his or her illness been addressed?'. Analysis

**Table 3.** Descriptive statistics and distribution for IPOS items at timepoint 1 (n = 376).

	Prevalence <sup>a</sup>	95% CI	Not at all (0)	Slight (1)	Moderate (2)	Severe (3)	Overwhelming/ all the time (4)	Missing
	%		%	%	%	%	%	%
<i>Physical symptoms</i>								
1 – Pain	62.3	57.4–67.2	17.8	18.6	30.3	26.1	5.9	1.3
2 – Shortness of breath	40.8	35.8–45.8	31.9	26.1	22.9	12.0	5.9	1.3
3 – Weakness or lack of energy	81.7	77.8–85.6	4.5	13.3	31.4	37.5	12.8	0.5
4 – Nausea	29.0	24.4–33.6	46.5	23.4	14.9	10.6	3.5	1.1
5 – Vomiting	14.6	11.0–18.2	73.1	10.1	7.2	6.1	1.3	2.1
6 – Poor appetite	48.9	43.9–53.9	27.4	20.2	22.6	18.9	7.4	3.5
7 – Constipation	42.2	37.2–47.2	39.9	16.5	19.1	16.5	6.6	1.3
8 – Sore or dry mouth	55.3	50.3–60.3	23.7	19.9	25.3	22.3	7.7	1.1
9 – Drowsiness	64.9	60.1–69.7	14.6	19.9	33.8	25.0	6.1	0.5
10 – Poor mobility	77.4	73.2–81.6	8.5	12.8	23.4	34.3	19.7	1.3
<i>Emotional symptoms</i>								
11 – Patient anxiety	71.0	66.4–75.6	13.6	14.4	29.5	25.3	16.2	1.1
12 – Family anxiety	84.8	81.2–88.4	6.9	5.9	17.0	33.2	34.6	2.4
13 – Depression	51.9	46.9–56.9	27.7	19.7	27.4	16.8	7.7	0.8
14 – Feeling at peace	72.1	67.6–76.6	8.8	17.0	18.9	34.8	18.4	2.1
<i>Communication/practical issues</i>								
15 – Sharing feelings	75.0	70.6–79.4	7.7	16.0	14.1	25.0	35.9	1.3
16 – Information	83.5	79.8–87.3	5.6	9.0	12.0	32.4	39.1	1.9
17 – Practical matters	28.7	24.1–33.3	42.8	26.3	16.0	6.6	6.1	2.1

IPOS: Integrated Palliative care Outcome Scale.

<sup>a</sup>Prevalence was defined as any IPOS symptoms/concerns specified as moderate, severe or overwhelming.

**Table 4.** Descriptive statistics and distribution for IPOS total and subscale scores at timepoint 1 (n = 376).

Total and sub- scale scores	# items	Range	Mean	SD	Skew	$\alpha^a$	Eigenvalue	% variance
IPOS Total Score	17	3–50	27.4	9.3	–.05	.77		
IPOS Physical symptoms	10	1–33	15.8	6.1	–.01	.70	3.5	24.9
IPOS Emotional symptoms	4	0–16	8.1	3.6	–.16	.68	1.7	12.3
IPOS Communication/ Practical Issues	3	0–12	3.4	2.7	.64	.58	1.2	8.3

IPOS: Integrated Palliative care Outcome Scale.

<sup>a</sup>Cronbach's alpha coefficient of internal reliability.

of the standard error of measurement for these items with low levels of agreement showed that errors for these items were close to 1 point on the IPOS.

The comparison of staff and patient ratings yielded similar results with acceptable to good agreement ( $\geq \kappa = 0.3$ ) achieved on 11 of 17 IPOS items with highest levels of agreement for 'Pain', 'Shortness of breath', 'Vomiting', and 'Constipation'. Again, items 'Having had enough information' ( $\kappa = 0.02$ ), 'Have practical matters been addressed?' ( $\kappa = 0.10$ ), and 'Sharing feelings with family or friends' ( $\kappa = 0.13$ ) showed low levels of agreement, together with the items 'Drowsiness' ( $\kappa = 0.11$ ), 'Feeling at peace' ( $\kappa = 0.26$ ) and 'Sore or dry mouth' ( $\kappa = 0.25$ ) (see Table 3). However, the proportion of scores that were within one score of a perfect match was still

high (above 70%) for these items, except for 'Drowsiness' (60.6%) and 'Sore or dry mouth' (65.1%). The proportion with agreement between patient and staff ratings was higher at the second assessment (see Appendix, Table 7).

### Responsiveness

Table 6 presents the mean changes in the IPOS total score in relation to patients' global report of change. SD at baseline for the total IPOS score was 9.2. Mean change scores for the total score were as large as 4.3 in the 'much improved' group and even larger (–9.6) for the group that described themselves as 'much worse'. However, associated standard deviations of change were comparably large, pointing towards potential misclassification

**Table 5.** Reliability assessment: weighted kappa ( $\kappa_w$ ) and proportion agreement within one score for test–retest reliability, inter-rater agreement between two independent staff ratings and inter-rater agreement between patient and staff ratings at timepoint 1.

IPOS Item	Test–retest			Inter-rater: Two staff			Inter-rater: patient-staff		
	N	$\kappa_{ww}$	%	n	$\kappa_{ww}$	%	n	$\kappa_{ww}$	%
Pain	66	0.49	81.8	92	0.72	91.3	348	0.59	87.1
Shortness of breath	66	0.78	92.4	91	0.82	92.3	345	0.62	86.1
Weakness or lack of energy	66	0.54	86.4	92	0.25	84.8	350	0.29	82.3
Nausea	66	0.75	94.0	91	0.63	94.5	346	0.46	81.2
Vomiting	66	0.62	89.4	90	0.61	92.2	342	0.54	88.3
Poor appetite	66	0.65	89.4	89	0.46	82.0	339	0.34	74.9
Constipation	66	0.69	89.4	86	0.41	80.2	342	0.47	77.5
Sore or dry mouth	66	0.79	90.9	85	0.49	84.7	343	0.25	65.1
Drowsiness	66	0.43	77.3	88	0.21	80.7	350	0.11	60.6
Poor mobility	66	0.71	86.4	91	0.46	83.5	348	0.42	74.4
Patient anxiety	66	0.64	80.3	95	0.44	87.4	347	0.35	75.2
Family anxiety	66	0.60	87.9	54	0.27	81.5	283	0.34	79.2
Depression	66	0.65	83.3	90	0.52	86.7	348	0.38	75.9
Feeling at peace	66	0.43	77.3	82	0.45	82.9	330	0.26	72.4
Sharing feelings	66	0.20	69.7	70	0.34	80.0	308	0.13	68.8
Information	66	0.39	75.8	77	0.14	70.1	332	0.02	70.2
Practical matters	66	0.55	83.3	59	0.20	77.9	317	0.10	68.5
Total IPOS	66	0.81	86.4	29	0.64	79.3	209	0.39	69.4
IPOS Physical Symptoms	66	0.76	89.4	94	0.57	81.9	355	0.39	72.1
IPOS Emotional Symptoms	66	0.67	80.3	95	0.45	64.2	351	0.38	64.4
IPOS Communication/Practical Issues	66	0.44	68.2	88	0.23	67.0	347	0.13	54.5

IPOS: Integrated Palliative care Outcome Scale.

**Table 6.** Mean total IPOS score changes (between T1–T2) by global change scale (a negative change scores indicates deterioration).

Total IPOS	n	Mean <sub>change</sub> $\pm$ SD <sub>change</sub>
Things have got...		
Much better	28	4.3 $\pm$ 6.1
A little better	90	3.0 $\pm$ 7.5
No change	55	1.7 $\pm$ 6.7
I don't know	3	-2.3 $\pm$ 13.5
A little worse	24	-0.3 $\pm$ 8.1
Much worse	5	-9.6 $\pm$ 8.0

IPOS: Integrated Palliative care Outcome Scale.

according to the global change criterion. Overall, a change of about 5 points in the total IPOS score represents a moderate effect size.

## Discussion

IPOS is a valid and reliable outcome measure for use with people with advanced illness. It has good structural validity, with three underlying factors – physical symptoms, emotional symptoms, and communication/practical issues. Unusually, this validation study included a high proportion of people with poor functional status, strengthening conclusions for the advanced illness population.

IPOS discriminates clearly between different palliative Phases of Illness.<sup>38,39</sup> The physical symptom subscale also discriminates between those with poor or high functional status. Almost all individual IPOS items show good agreement when re-tested in stable patients. There is acceptable or good agreement between most patient self-reported and staff proxy-reported items. Most importantly, the total IPOS score showed a change in keeping with patient-report of the overall change in their symptoms and other concerns, both in direction and size of change.

The changing age distribution of the population and increasing prevalence of multi-morbidities (with more complex health needs)<sup>55</sup> require outcome measures to work across conditions and in advanced illness.<sup>56</sup> Only measures that reflect patient priorities can support a truly patient-centred approach to care.<sup>57</sup> Until now, outcome measures extending beyond symptoms or quality-of-life for this population have been lacking. Health-related quality-of-life measures, often heavily based on physical function, show low sensitivity with large floor/ceiling effects among those with advanced, multi-morbid disease and do not capture the main priorities of those affected.<sup>58–64</sup>

The IPOS has features which set it apart from other outcome measures commonly used in the context of advanced disease. Including how symptoms or other concerns have affected the individual themselves is a distinct characteristic not commonly sought in quality-of-life or symptom



tools.<sup>65</sup> The ESAS,<sup>31</sup> the M.D. Anderson Symptom Inventory (MDASI),<sup>66</sup> the Symptom Distress Scale<sup>67</sup> and the Palliative Problem Severity Score (PCPSS)<sup>38</sup> all focus on *severity* of symptoms. With the exception of PCPSS, they score physical and psychological symptoms, excluding concerns such as family issues (family anxiety, sharing feelings with family or friends), spirituality, practical issues, information needs, and communication concerns.<sup>58–64</sup>

While existing tools are well-validated,<sup>41,67–71</sup> proxy-reported versions are much less well established.<sup>72,73</sup> IPOS may also capture the impact of symptoms and concerns differently. In terms of overall validity, the performance of the new, refined IPOS is comparable to both the original POS and similar measures in the field. In the original POS validation,<sup>12</sup> mid-range correlations to the EORTC QLQ-C30 physical, non-physical and quality-of-life subscales were reported. Mid-range correlations were also apparent in comparison of POS with the Rotterdam Symptom Checklist,<sup>74</sup> the Italian POS with the FACIT-SP and EORTC QLQ-C15-Pal,<sup>55</sup> the POS-S with the Rotterdam checklist, the MDASI and EORTC measures,<sup>12,55,74</sup> and in this validation study when comparing IPOS to FACT-G and ESAS. This result – of mid-range rather than high correlations – is likely because existing scales largely focus on the *severity* of symptoms, whereas POS and IPOS focus on how a person is *affected*. In line with this, comparison of the POS pain item with the Brief Pain Inventory's pain impairment item yielded a higher correlation than to the Pain Severity item.<sup>74</sup> The mid-range correlations ( $\leq 0.50$ ) between aspects of psychological well-being across questionnaires further demonstrate the different dimensions included in IPOS, covering wider issues of spiritual and family well-being. The consistently low correlations of the communication/practical items with other outcome measures across studies point towards the uniqueness of this aspect.<sup>12,55,74,75</sup>

The second distinct feature of this validation of IPOS is the broader testing of reliability. In terms of test–retest reliability, we found mostly good to very good agreement (weighted kappa values 0.50–0.80). These values are higher than in similar studies of test–retest reliability of either POS<sup>12,55,74,75</sup> or ESAS,<sup>31,70,76</sup> perhaps explained by using an external criterion to judge stability, rather than assuming stability over 24–48 h, an assumption that might not be justified in a fast-changing palliative population.<sup>58</sup> However, some items of IPOS ('Sharing feelings with family or friends', 'Having enough information', 'Feeling at peace', and 'Drowsiness') showed less agreement. These items also showed low agreement in the comparison of patient and staff ratings. This is consistent with prior studies of the biases affecting proxy assessments,<sup>77</sup> which suggest systematic overestimation of physical symptoms and underestimation of psychological well-being and information needs.<sup>78–81</sup> The low agreement for the information item had also been observed in studies of POS; Higginson and Gao<sup>82</sup> reported a weighted kappa value as low as 0.04

for this item, and results by Dawber et al.<sup>78</sup> and Van Soest-Poortvliet et al.<sup>83</sup> are similar. A study of the Palliative care Problem Severity Index<sup>38</sup> identified features of the raters (e.g. new staff member with new patient), patient characteristics (e.g. communication problems, drowsiness), or family characteristics (e.g. lacking interaction with family), as impeding agreement. This result has also been observed in a study looking at proxy ratings of the McGill Quality of Life questionnaire.<sup>79</sup> These features may also have been present in the IPOS validation study. Fluctuating symptoms, in particular drowsiness, may also contribute to this as demonstrated for the comparable ESAS item.<sup>84,85</sup>

This validation was cross-cultural, conducted in two countries simultaneously, and we believe this strengthened both cognitive development of the measure<sup>24</sup> and this full-scale validation. Despite some skewness in distributions of individual items, both parametric and non-parametric statistics showed highly similar and robust results, both in terms of correlations and test of differences. However, there were limitations in the population studied; only 15% of patient participants had non-cancer conditions. IPOS needs to be further tested in non-cancer conditions, and refinements for different diseases may be required. Indeed, development of a version for use in cognitive impairment and dementia is already well under way.<sup>86</sup>

The use of consecutive enrolment ensured IPOS was validated in a broadly clinically representative group. However, selection bias cannot be ruled out. This validation also mirrored conditions for IPOS use in clinical practice – the 'least controlled' use – for instance with absence of specific staff training prior to implementation.<sup>87</sup>

Despite incorporation of a global criterion for change, it was not possible to derive values for a minimal clinically important change for improvement or deterioration for the total IPOS, its subscales, and individual items. Such a feature is desirable, as optimal cut-offs for individual symptoms in particular can trigger specific clinical actions, such as referral to a palliative care team, help triage patients within services<sup>88</sup> and therefore extend the clinical and research utility of an outcome measure.<sup>72,89,90</sup>

### *Clinical and research implications*

This study has demonstrated IPOS is valid and reliable. Because it is brief and underpinned by the symptoms and concerns of people with advanced illness, it will be invaluable for clinical practice and research. To implement such a measure into routine clinical practice needs training. A recent survey of the use of ESAS showed a range of training needs and other barriers.<sup>85</sup> We are already working on best ways to implement, using the national Outcomes Assessment and Complexity Collaborative in the United Kingdom. This is based on the well-established Australian Palliative Care Outcomes Collaborative.<sup>91</sup>

## Conclusion

The IPOS is a valid and reliable outcome measure for use with people with advanced illness, both in its patient self-report and staff proxy-report versions. It is suitable for assessing and monitoring symptoms and concerns in advanced illness, monitoring change over time, determining the impact of healthcare interventions, and demonstrating quality of care. This will be invaluable for clinical care, audit and research, and represents a major step forward internationally for outcome measurement in advanced illness.

## Acknowledgements

Our sincere thanks to all participants and to Fern Brookes for proof-reading and formatting. Especial thanks are due to our patient and public involvement group, without which this work would have been much less relevant for people with advanced illness.

## Authors' contributions

I.J.H. developed and validated the original POS. C.B. translated, culturally adapted and validated the German POS based on the original English POS. F.M., C.B. and I.J.H. developed the IPOS, conceived this validation study, and led the protocol for this study. F.M. and I.J.H. led the grant application for funding. E.I.G., A.F., E.S., C.B. and F.M. drafted and submitted ethical and other approvals, with input from I.J.H., E.I.G., A.F., N.L., E.S., and S.T.S. J.D., F.B., B.O., F.H., and S.S. provided critical input into the protocol and undertook data collection. C.R. undertook the main analyses with critical input from I.J.H., F.M. and C.B. F.M. and C.R. drafted the paper, and all authors read, refined, and provided critical input into the revisions of the manuscript, and approved the final manuscript. C.B. and I.J.H. are equal co-senior authors.

## Availability of data

The datasets generated and/or analysed during the current study are not publicly available to protect confidentiality but aggregated data are available from the corresponding author on reasonable request. The IPOS and other POS measures are freely available – see [www.pos-pal.org](http://www.pos-pal.org).<sup>92</sup>

## Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: I.J.H. developed and validated the original POS. C.B. translated, culturally adapted and validated the German version of POS. Otherwise, the authors declare that they have no competing interests.

## Ethics and consent

Ethical approval for this study was granted in the United Kingdom by the Dulwich National Research Ethics Committee, London, UK (Reference Number 124991) and in Germany by the Local Research Ethics Committee of Munich University (No.

169-13) and University Hospital Cologne (No.14-198). All participants gave written consent, including consent for publication of anonymised findings, and formal Distress Protocols were adhered to.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is funded by the National Institute for Health Research, through a Programme Grant for Applied Research (the C-CHANGE project: RP-PG-1210-12015). The views and opinions expressed by authors do not necessarily reflect those of the National Health Service, the National Institute for Health Research, MRC, CCF, NETSCC, the National Institute for Health Research Programme Grants for Applied Research programme or the Department of Health and Social Care.





## Supplemental material

Supplemental material for this article is available online.

## Title of measure

The Palliative care Outcome Scale can also be called The Patient Outcome Scale, when this is more appropriate to the context for use.

## ORCID iDs

Fliess EM Murtagh  <https://orcid.org/0000-0003-1289-3726>  
 Christina Ramsenthaler  <https://orcid.org/0000-0002-9996-1818>  
 Alice Firth  <https://orcid.org/0000-0003-0726-0502>  
 Natasha Lovell  <https://orcid.org/0000-0001-6594-799X>  
 Eva Schildmann  <https://orcid.org/0000-0001-9756-9954>

## References

1. World Health Organization. Strengthening of palliative care as a component of integrated treatment throughout the life course. *J Pain Palliat Care Pharmacother* 2014; 28(2): 130–134.
2. Christensen K, Doblhammer G, Rau R, et al. Ageing populations: the challenges ahead. *Lancet* 2009; 374(9696): 1196–1208.
3. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015; 385(9963): 117–171.
4. Porter ME and Teisberg EO. *Redefining health care: creating value-based competition on results*. Boston, MA: Harvard Business School Press, 2006.
5. Epping-Jordan JE, Pruitt SD, Bengoa R, et al. Improving the quality of health care for chronic conditions. *Qual Saf Health Care* 2004; 13(4): 299–305.
6. Dawson J, Doll H, Fitzpatrick R, et al. The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010; 340: c186.

7. Hays RD, Spritzer KL, Fries JF, et al. Responsiveness and minimally important difference for the patient-reported outcomes measurement information system (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. *Ann Rheum Dis* 2015; 74(1): 104–107.
8. Gilbert A, Sebag-Montefiore D, Davidson S, et al. Use of patient-reported outcomes to measure symptoms and health related quality of life in the clinic. *Gynecol Oncol* 2015; 136(3): 429–439.
9. Dur M, Sadlonova M, Haider S, et al. Health determining concepts important to people with Crohn's disease and their coverage by patient-reported outcomes of health and wellbeing. *J Crohns Colitis* 2014; 8(1): 45–55.
10. Berg SK, Svanholm J, Lauberg A, et al. Patient-reported outcomes at hospital discharge from Heart Centres, a national cross-sectional survey with a register-based follow-up: the DenHeart study protocol. *BMJ Open* 2014; 4(5): e004709.
11. Mak KS, van Bommel AC, Stowell C, et al. Defining a standard set of patient-centred outcomes for lung cancer. *Eur Respir J* 2016; 48(3): 852–860.
12. Hearn J and Higginson IJ. Development and validation of a core outcome measure for palliative care: the palliative care outcome scale. *Qual Health Care* 1999; 8(4): 219–227.
13. Dzingina M, Higginson IJ, McCrone P, et al. Development of a patient-reported palliative care-specific health classification system: the POS-E. *Patient* 2017; 10(3): 353–365.
14. Higginson IJ and Donaldson N. Relationship between three palliative care outcome scales. *Health Qual Life Outcomes* 2004; 2: 68.
15. Siegert RJ, Gao W, Walkey FH, et al. Psychological wellbeing and quality of care: a factor-analytic examination of the palliative care outcome scale. *J Pain Symptom Manage* 2010; 40(1): 67–74.
16. Harding R, Selman L, Simms VM, et al. How to analyze palliative care outcome data for patients in Sub-Saharan Africa: an international, multicenter, factor analytic examination of the APCA African POS. *J Pain Symptom Manage* 2013; 45(4): 746–752.
17. Harding R, Selman L, Agupio G, et al. Validation of a core outcome measure for palliative care in Africa: the APCA African Palliative Outcome Scale. *Health Qual Life Outcomes* 2010; 8: 10.
18. Bausewein C, Le Grice C, Simon S, et al. 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): 304–313.
19. Collins ES, Witt J, Bausewein C, et al. A systematic review of the use of the Palliative Care Outcome Scale and the Support Team Assessment Schedule in Palliative Care. *J Pain Symptom Manage* 2015; 50(6): 842–853.e19.
20. Aspinall F, Hughes R and Higginson IJ. *A user's guide to the Palliative care Outcome Scale*. London: Palliative Care & Policy Publications, 2002.
21. Solano JP, Gomes B and Higginson IJ. A comparison of symptom prevalence in far advanced cancer, AIDS, heart disease, chronic obstructive pulmonary disease and renal disease. *J Pain Symptom Manage* 2006; 31(1): 58–69.
22. Higginson IJ, Simon ST, Benalia H, et al. Which questions of two commonly used multidimensional palliative care patient reported outcome measures are most useful? Results from the European and African PRISMA survey. *BMJ Support Palliat Care* 2012; 2(1): 36–42.
23. Harding R, Simon ST, Benalia H, et al. The PRISMA Symposium 1: outcome tool use. Disharmony in European outcomes research for palliative and advanced disease care: too many tools in practice. *J Pain Symptom Manage* 2011; 42(4): 493–500.
24. Schildmann EK, Groeneveld EI, Denzel J, et al. Discovering the hidden benefits of cognitive interviewing in two languages: the first phase of a validation study of the Integrated Palliative care Outcome Scale. *Palliat Med* 2016; 30(6): 599–610.
25. Terwee CB, Bot SD, deBoer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60(1): 34–42.
26. Abernethy AP, Shelby-James T, Fazekas BS, et al. The Australian modified Karnofsky Performance Status (AKPS) scale: a revised scale for contemporary palliative care clinical practice. *BMC Palliat Care* 2005; 4(1): 7.
27. Murphy EL, Murtagh FE, Carey I, et al. Understanding symptoms in patients with advanced chronic kidney disease managed without dialysis: use of a short patient-completed assessment tool. *Nephron Clin Pract* 2009; 111(1): c74–c80.
28. Bausewein C, Fegg M, Radbruch L, et al. Validation and clinical application of the German version of the palliative care outcome scale. *J Pain Symptom Manage* 2005; 30(1): 51–62.
29. Eisenchlas JH, Harding R, Daud ML, et al. Use of the palliative outcome scale in Argentina: a cross-cultural adaptation and validation study. *J Pain Symptom Manage* 2008; 35(2): 188–202.
30. Bruera E, Kuehn N, Miller MJ, et al. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. *J Palliat Care* 1991; 7(2): 6–9.
31. Chang VT, Hwang SS and Feuerman M. Validation of the Edmonton Symptom Assessment Scale. *Cancer* 2000; 88(9): 2164–2171.
32. Watanabe SMNC, Beaumont C, Johnson L, et al. A multi-centre comparison of two numerical versions of the Edmonton Symptom Assessment System in palliative care patients. *J Pain Symptom Manage* 2011; 41: 456–468.
33. Cella D, Tulskey DS, Gray G, et al. The Functional Assessment of Cancer Therapy Scale: development and validation of the general measure. *J Clin Oncol* 1993; 11(3): 570–579.
34. Jaeschke R, Singer J and Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10(4): 407–415.
35. Higginson I. Audit methods: validation and in-patient use. In: Higginson I (ed.) *Clinical audit in palliative care*. Oxford: Radcliffe Medical Press, 1993, pp. 48–54.
36. Higginson I. *The development, validity, reliability and practicality of a new measure of palliative care: the Support Team Assessment Schedule*. PhD thesis, University College London, London, 1992, <http://discovery.ucl.ac.uk/1317889/1/296225.pdf> (accessed 23 June 2015).
37. Higginson I and McCarthy M. Validity of the Support Team Assessment Schedule: do staffs' ratings reflect those made by patients or their families? *Palliat Med* 1993; 7: 219–228.
38. Masso M, Allingham SF, Banfield M, et al. Palliative Care Phase: inter-rater reliability and acceptability in a national study. *Palliat Med* 2015; 29(1): 22–30.

39. Mather H, Guo P, Firth A, et al. Phase of illness in palliative care: cross-sectional analysis of clinical data from community, hospital and hospice patients. *Palliat Med* 2018; 32(2): 404–412.
40. Guyatt GH, Norman GR, Juniper EF, et al. A critical look at transition ratings. *J Clin Epidemiol* 2002; 55(9): 900–908.
41. Higginson IJ and McCarthy M. Validity of the support team assessment schedule: do staffs' ratings reflect those made by patients or their families? *Palliat Med* 1993; 7(3): 219–228.
42. Satorra A and Bentler PM. Corrections to test-statistics and standard errors in covariance structure analysis. In: von Eye A and Clogg CC (eds) *Latent variable analysis: applications for developmental research*. Thousand Oaks, CA: SAGE, pp. 399–419.
43. Bentler PM. Comparative fit indices in structural models. *Psychological Bulletin*; 107: 238–246.
44. Brown TA. *Confirmatory factor analysis for applied research*. New York and London: The Guilford Press, 2006.
45. Osborne TR, Ramsenthaler C, Schey SA, et al. Improving the assessment of quality of life in the clinical care of myeloma patients: the development and validation of the Myeloma Patient Outcome Scale (MyPOS). *BMC Cancer* 2015; 15: 280.
46. Fayers PM and Machin D. Scores and measurements: validity, reliability, sensitivity. In: Fayers P (ed.) *Quality of life: the assessment, analysis and interpretation of patient reported outcomes*. 2nd ed. Chichester: John Wiley & Sons, 2007, pp. 77–108.
47. de Vet HCW, Terwee CB, Mokkink LB, et al. Field testing: item reduction and data structure. In: de Vet HCW, Terwee CB, Mokkink LB, et al. (eds) *Measurement in medicine*. Cambridge: Cambridge University Press, 2011, pp. 65–95.
48. Landis R and Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33(1): 159–174.
49. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; 76(5): 378–382.
50. Crosby RD, Kolotkin RL and Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; 56(5): 395–407.
51. SPSS Inc. *SPSS for Windows, Version 24.0*. Armonk, NY: IBM Corporation, 2016.
52. Rosseel Y. lavaan: an R package for structural equation modelling. *J Stat Software* 2012; 48(2): 1–36.
53. Bentler PM and Chou C. Practical issues in structural modelling. *Sociol Method Res* 1987; 16: 78–117.
54. Pornprasertmanit S, Miller P, Schoemann A, et al. simsem: simulated structural equation modeling. *R package* 2016, <http://www.cran.r-project.org> (accessed 1 May 2018).
55. Costantini M, Rabitti E, Beccaro M, et al. Validity, reliability and responsiveness to change of the Italian palliative care outcome scale: a multicentre study of advanced cancer patients. *BMC Palliat Care* 2016; 15: 23.
56. Currow DC, Allingham S, Yates P, et al. Improving national hospice/palliative care service symptom outcomes systematically through point-of-care data collection, structured feedback and benchmarking. *Support Care Cancer* 2015; 23(2): 307–315.
57. Higginson IJ, Hart S, Koffman J, et al. Needs assessment in palliative care: an appraisal of definitions and approaches used. *J Pain Symptom Manage* 2007; 33(5): 500–505.
58. Higginson IJ and Carr AJ. Measuring quality of life: using quality of life measures in the clinical setting. *BMJ* 2001; 322(7297): 1297–1300.
59. Heyland DK, Dodek P, Rocker G, et al. What matters most in end-of-life care: perceptions of seriously ill patients and their family members. *CMAJ* 2006; 174(5): 627–633.
60. Lynn J, Teno JM, Phillips RS, et al. Perceptions by family members of the dying experience of older and seriously ill patients. *Ann Intern Med* 1997; 126(2): 97–106.
61. Singer PA, Martin DK and Kelner M. Quality end-of-life care: patients' perspectives. *JAMA* 1999; 281(2): 163–168.
62. Steinhilber KE, Clipp EC, McNeilly M, et al. In search of a good death: observations of patients, families, and providers. *Ann Intern Med* 2000; 132(10): 825–832.
63. Teno JM, Casey VA, Welch LC, et al. Patient-focused, family-centered end-of-life medical care: views of the guidelines and bereaved family members. *J Pain Symptom Manage* 2001; 22(3): 738–751.
64. Teno JM, Clarridge BR, Casey V, et al. Family perspectives on end-of-life care at the last place of care. *JAMA* 2004; 291(1): 88–93.
65. Osborne TR, Ramsenthaler C, Siegert RJ, et al. What issues matter most to people with multiple myeloma and how well are we measuring them? A systematic review of quality of life tools. *Eur J Haematol* 2012; 89(6): 437–457.
66. Cleland CS, Mendoza TR, Wang XS, et al. Assessing symptom distress in cancer patients: the M.D. *Cancer* 2000; 89(7): 1634–1646.
67. McCorkle R and Young K. Development of a symptom distress scale. *Cancer Nurs* 1978; 1(5): 373–378.
68. Miyasaki JM, Long J, Mancini D, et al. Palliative care for advanced Parkinson disease: an interdisciplinary clinic and new scale, the ESAS-PD. *Parkinsonism Relat Disord* 2012; 18(Suppl 3): S6–S9.
69. Davison SN, Jhangri GS and Johnson JA. Longitudinal validation of a modified Edmonton Symptom Assessment System (ESAS) in haemodialysis patients. *Nephrol Dial Transplant* 2006; 21(11): 3189–3195.
70. Aoun SM, Monterosso L, Kristjanson LJ, et al. Measuring symptom distress in palliative care: psychometric properties of the Symptom Assessment Scale (SAS). *J Palliat Med* 2011; 14(3): 315–321.
71. Aktas A, Walsh D and Kirkova J. The psychometric properties of cancer multisymptom assessment instruments: a clinical review. *Support Care Cancer* 2015; 23(7): 2189–2202.
72. Hui D and Bruera E. The Edmonton Symptom Assessment System 25 years later: past, present, and future developments. *J Pain Symptom Manage* 2017; 53(3): 630–643.
73. Nekolaichuk C, Watanabe S and Beaumont C. The Edmonton Symptom Assessment System: a 15-year retrospective review of validation studies (1991–2006). *Palliat Med* 2008; 22(2): 111–122.
74. Pelayo-Alvarez M, Perez-Hoyos S and Agra-Varela Y. Reliability and concurrent validity of the Palliative Outcome Scale, the Rotterdam Symptom Checklist, and the Brief Pain Inventory. *J Palliat Med* 2013; 16(8): 867–874.
75. Sleeman KE and Higginson IJ. A psychometric validation of two brief measures to assess palliative need in patients severely affected by multiple sclerosis. *J Pain Symptom Manage* 2013; 46(3): 406–412.

76. Dong Y, Chen H, Zheng Y, et al. Psychometric validation of the Edmonton Symptom Assessment System in Chinese patients. *J Pain Symptom Manage* 2015; 50(5): 712–717.e2.
77. McPherson CL and Addington-Hall JM. Judging the quality of care at the end of life: can proxies provide reliable information? *Soc Sci Med* 2003; 56(1): 95–109.
78. Dawber R, Armour K, Ferry P, et al. Comparison of informal caregiver and named nurse assessment of symptoms in elderly patients dying in hospital using the palliative outcome scale. *BMJ Support Palliat Care*. Epub ahead of print 12 January 2016. doi:10.1136/bmjspcare-2015-00850
79. Jones JM, McPherson CJ, Zimmermann C, et al. Assessing agreement between terminally ill cancer patients' reports of their quality of life and family caregiver and palliative care physician proxy ratings. *J Pain Symptom Manage* 2011; 42(3): 354–365.
80. Crocker TF, Smith JK and Skevington SM. Family and professionals underestimate quality of life across diverse cultures and health conditions: systematic review. *J Clin Epidemiol* 2015; 68(5): 584–595.
81. Roydhouse JK and Wilson IB. Systematic review of caregiver responses for patient health-related quality of life in adult cancer care. *Qual Life Res* 2017; 26(8): 1925–1954.
82. Higginson IJ and Gao W. Caregiver assessment of patients with advanced cancer: concordance with patients, effect of burden and positivity. *Health Qual Life Outcomes* 2008; 6: 42.
83. Van Soest-Poortvliet MC, vanderSteen JT, Zimmerman S, et al. Psychometric properties of instruments to measure the quality of end-of-life care and dying for long-term care residents with dementia. *Qual Life Res* 2012; 21(4): 671–684.
84. Bergh I, Kvaalem IL, Aass N, et al. What does the answer mean? A qualitative study of how palliative cancer patients interpret and respond to the Edmonton Symptom Assessment System. *Palliat Med* 2011; 25(7): 716–724.
85. CarliButtenschoen D, Stephan J, Watanabe S, et al. Health care providers' use and knowledge of the Edmonton Symptom Assessment System (ESAS): is there a need to improve information and training? *Support Care Cancer* 2014; 22(1): 201–208.
86. Ellis-Smith C, Evans CJ, Murtagh FE, et al. Development of a caregiver-reported measure to support systematic assessment of people with dementia in long-term care: the Integrated Palliative care Outcome Scale for Dementia. *Palliat Med* 2017; 31(7): 651–660.
87. Evans CJ, Benalia H, Preston NJ, et al. The selection and use of outcome measures in palliative and end-of-life care research: the MORECare international consensus workshop. *J Pain Symptom Manage* 2013; 46(6): 925–937.
88. Dhaliwal S, Salins N, Deodhar J, et al. Pilot testing of triage coding system in home-based palliative care using Edmonton Symptom Assessment Scale. *Indian J Palliat Care* 2016; 22(1): 19–24.
89. Hui D, Shamieh O, Paiva CE, et al. Minimal clinically important differences in the Edmonton Symptom Assessment Scale in cancer patients: a prospective study. *Cancer* 2015; 121: 3027–3035.
90. Hui D, Shamieh O, Paiva CE, et al. Minimal clinically important difference in the physical, emotional, and total symptom distress scores of the Edmonton Symptom Assessment System. *J Pain Symptom Manage* 2016; 51(2): 262–269.
91. Australian Palliative Care Outcomes Collaborative. Australian Health Services Research Institute, University of Wollongong, Australia, 2007, <https://ahsri.uow.edu.au/pcoc/index.html> (accessed 15 October 2018).
92. POS family of measures. Cicely Saunders Institute, King's College London, 2018, [www.pos-pal.org](http://www.pos-pal.org) (accessed 15 October 2018).