Evaluation of quality-of-life measures for use in palliative care: a systematic review

Gwenda Albers, Michael A Echteld, Henrica CW de Vet, Bregje D Onwuteaka-Philipsen, Mecheline HM van der Linden and Luc Deliens

Palliat Med 2010; 24; 17 originally published online Oct 20, 2009;
DOI: 10.1177/0269216309346593

Original Article

PALLIATIVE MEDICINE

Palliative Medicine
24(1) 17–37

DOI: 10.1177/0269216309346593
pmj.sagepub.com

SAGE

Evaluation of quality-of-life measures for use in palliative care: a systematic review

Gwenda Albers *Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands; Palliative Care Centre of Expertise, VU University Medical Center, Amsterdam, The Netherlands*

Michael A Echteld *Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands; Palliative Care Centre of Expertise, VU University Medical Center, Amsterdam, The Netherlands*

Henrica CW de Vet *Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands*

Bregje D Onwuteaka-Philipsen *Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands; Palliative Care Centre of Expertise, VU University Medical Center, Amsterdam, The Netherlands*

Mecheline HM van der Linden *Department of Medical Psychology and Department of Medical Oncology, VU University Medical Center, Amsterdam, The Netherlands; Palliative Care Centre of Expertise, VU University Medical Center, Amsterdam, The Netherlands*

Luc Deliens *Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands; Palliative Care Centre of Expertise, VU University Medical Center, Amsterdam, The Netherlands; End-of-Life Care Research Group, Vrije Universiteit Brussel, Brussels, Belgium*

Abstract

Purpose: In this literature review we evaluated the feasibility and clinimetric quality of quality-of-life (QoL) measurement instruments suitable for use in palliative care. **Methods:** We conducted a systematic literature review to identify instruments measuring (at least one domain of) QoL. We selected articles that present data on patients receiving palliative care and at least one measurement property. A checklist was used to describe the characteristics of the instruments, and a widely accepted rating list was used to evaluate the clinimetric aspects. **Results:** 29 instruments were identified and evaluated, most of which were targeted at palliative patients in general. None of the instruments demonstrated satisfactory results for all measurement properties. Fourteen instruments received positive ratings for construct validity. Thirteen instruments were tested for reliability, but only two were tested adequately and had positive results (ICC>0.70). Responsiveness was not tested adequately for any of the instruments. Very few of the studies provided information on the interpretation of the scores. Overall, the MQOL, followed by the QUAL-E and the QODD, received the best ratings for their measurement properties. **Conclusions:** Many measurement instruments were identified, but most had not yet been adequately evaluated. The evaluation of existing instruments with good content validity should have priority over the development of new instruments.

Keywords

palliative care, end of life, quality of life, instruments

Introduction

The interest in palliative care has significantly increased in the past decade. The main focus of palliative care is to improve the quality of life of patients and their families who face the problems associated with a life-threatening illness.[1] Palliative care may entail any form of medical care or treatment that concentrates on the prevention and relief of suffering. Any combination of pain and symptom management, psychological care and spiritual care, and social support can be applied to improve the quality of life of patients for whom there are no longer any curative treatment options.[2] Palliative care is most commonly associated with

Corresponding author:
Department of Public and Occupational Health, EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands; Palliative Care Centre of Expertise, VU University Medical Center, Amsterdam, The Netherlands. Email: g.albers@vumc.nl

cancer patients, but it can be applied to all patients with incurable diseases, for example patients with heart failure, renal disease or neurodegenerative diseases such as multiple sclerosis or amyotrophic lateral sclerosis.

Quality-of-life measurement is an important aspect of palliative care, given that maximizing the quality-of-life of terminally ill patients is the main aim of this type of care. A large variety of quality-of-life measurement instruments are appropriate for use in palliative care. However, both feasibility (for example, the number of questions and the completion time needed) and clinimetric quality varies widely over these instruments. Furthermore, at present there is no agreement on how quality-of-life should be measured, or which is the best instrument to use. Consequently, many different quality-of-life questionnaires are used, and new ones continue to be developed. We felt the need to determine which are adequate instruments, in order to facilitate decision making with regard to the most appropriate instruments for use in research or clinical practice.

A variety of earlier reviews have identified quality-of-life measurement instruments that are appropriate for use in palliative care.[3–10] However, none of these reviews could serve as a guide for the adequate and comprehensive choice of a questionnaire for research or clinical practice. First of all, because many reviews[4,7,9] have focused on instruments that have been specifically designed for cancer patients, whereas quality-of-life measurement in patients with other terminal diseases is also of great significance. Furthermore, Jordhoy et al.[5] recently published a review of quality-of-life measures, but they focused on the aspect of physical functioning only. Mularski et al.[10] reviewed not only quality-of-life instruments, but all measures of end-of-life care, including instruments to measure satisfaction and the quality of the care, caregiver well-being, grief and bereavement. Additionally, most reviews could possibly have missed some studies which focused on domain-specific instruments, because the reviewers searched for instruments measuring overall quality of life. In particular, spirituality-specific instruments could have been missed, because spirituality has only recently been considered to be important for the quality of life of terminally ill patients.[11,12] As a consequence, spirituality is somewhat under represented in several quality-of-life measurement instruments. Moreover, all of the reviews[3–10] described the content and measurement properties of the instruments, but none had a rating list with explicit criteria assessing measurement properties. Therefore, it remains difficult to compare the quality of various measurement instruments, and to determine what a good, or the best, questionnaire is, given any combination of measurement purpose and patient group.

The purpose of the present study was to make an inventory of all currently available quality-of-life measurement instruments that are suitable for the use in palliative care and to assess the content and clinimetric quality of these instruments. This can help investigators and clinicians in their choice of an adequate measurement instrument that is applicable in palliative care.

Methods

Selection of the measurement instruments

We searched PubMed, Embase, CINAHL and PsycINFO for relevant literature in the English and the Dutch language (January 1990 to April 2008). The following keywords were used to identify eligible studies: palliative care, terminal care, hospice care, end-of-life, and quality-of-life (MESH term or text word), combined with a search filter for clinimetric studies. Because spirituality is somewhat under represented in a number of quality-of-life instruments, we added two search terms: 'religion and psychology' (MESH term) and 'spiritual' (text word). Appendix 1 presents a detailed overview of the search strategy. All abstracts were reviewed by one reviewer to assess whether the study was eligible for inclusion in the review. We applied the following inclusion criteria: (1) the study should describe the development or validation of a measurement tool; (2) the measurement instrument should measure (at least one domain of) quality of life in a population of patients for whom there are no further curative treatment options; (3) the study should have investigated at least one measurement property of the instrument; (4) the measurement instrument should have been validated in an English or a Dutch population. We excluded studies concerning instruments that are intended to measure the quality of and/or satisfaction with palliative care. Studies published as a clinical trial, case-report, editorial, bibliography or review were also excluded. If there was any uncertainty about inclusion, eligibility was assessed by two reviewers based on the full text of the article.

Data extraction

Data were extracted from the articles for the description of the instrument characteristics and the quality assessment by two independent reviewers (GA and one of the other authors). The results of the data extraction and the ratings for the clinimetric characteristics were compared, and any disagreements between the reviewers were discussed and resolved in consensus meetings. If necessary, any remaining disagreement was resolved by a third reviewer (HCWdV or MAE).

The quality assessment ratings were based on the quality criteria for measurement properties defined by Terwee et al. [13] and the preliminary version of the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN).[14]

Instrument characteristics: The descriptive data extracted from the studies included: (a) the target population; (b) the quality-of-life domains the instrument is intended to measure; (c) the number of items; (d) the number of response options; (e) the scoring algorithm (for example, sub scale scores and/or total score); (f) the recall period; (g) the time needed to complete a questionnaire; (h) the mode of administration (for example, (proxy) self-report or interview); and (i) whether the full text of the instrument is available. These aspects describe the design, content and application of measurement instruments, and provide clinicians and researchers with information which could help them to decide which instruments may be appropriate and/or feasible for a particular study or setting.

Measurement properties: Measurement properties convey information about the clinimetric quality of a measurement instrument, and can guide researchers and clinicians in making a choice between various potentially appropriate instruments. We rated content and construct validity, internal consistency, reliability, responsiveness and interpretability. The quality criteria will be described in more detail below (see also Appendix 2).

Validity: Validity refers to the extent to which an instrument measures what it is intended to measure.[15] The instruments were evaluated for both content and construct validity. Content validity refers to the degree to which the domains of interest are represented by the items in the questionnaire.[16] These items must reflect aspects that are important to patients for whom there are no further curative treatment options. Therefore, the involvement of patients in the item selection is a requirement, in combination with reference to the literature or consultation with experts. There should be a clear description of the measurement aims, the target population and the item selection. Lastly, the full text of the instrument must be available to achieve a positive rating.

Construct validity refers to the extent to which the scores for a particular instrument correspond to other measures in a manner that is consistent with theoretical expectations concerning the constructs that are measured.[17] Construct validity should be assessed by testing predefined hypotheses (for example, about expected correlations between (scales of) a questionnaire and another comparable instrument). A positive rating is achieved if the hypotheses are specified in advance and at least 75% of the hypotheses are confirmed.

Internal consistency: Internal consistency is a measure of the extent to which items in a questionnaire (sub)scale are correlated, thus measuring the same construct. Factor analysis should be applied to determine the homogeneity of items in a (sub)scale. To determine the internal consistency Cronbach's alpha should be calculated for each (sub)scale separately. A positive rating is achieved when factor analysis is performed in an adequate study size (7* number of items AND >100) and Cronbach's alpha for each sub scale is between 0.70 and 0.90. Note that Cronbach's alpha is only relevant if the instrument is based on a reflective model. In a reflective model, the construct to be measured is reflected in the items, in contrast to a formative model, in which the items are causal and form the construct to be measured.[18]

Reliability: Reliability concerns the degree to which repeated measurements in stable persons provide similar results. The time-interval between two measurements needs to be short enough to ensure that no change in quality of life has occurred and long enough to prevent recall bias. A time-interval of 1 week was considered to be appropriate for terminally ill patients. We assessed the test–retest reliability and the absolute measurement error. Reliability refers to the extent to which the instrument is able to distinguish patients from each other, despite measurement error. Reliability was assessed as positive if an intraclass correlation coefficient (ICC) or Kappa of at least 0.70 was calculated for each domain.

Absolute measurement error, measuring lack of agreement, estimates the absolute difference between two repeated measurements, and is expressed in the dimension of measurement. The standard error of measurement (SEM), or the smallest detectable change (SDC) are adequate measures of absolute measurement error. The SDC must be smaller than the minimal important change (MIC), or the MIC must be outside the limits of agreement (LOA) to score a positive rating. Because the MIC value is a relatively new approach, and not yet widely known, a positive rating is also given if the authors have provided convincing arguments that the measurement error was acceptable. In both the evaluation of test–retest reliability and measurement error, the sample size must be at least 50 patients.

Responsiveness: Responsiveness refers to the ability of an instrument to detect important change over time in the concept being measured.[19] The evaluation of

responsiveness requires predictions about how the results of the questionnaire should correlate with other related measurements. Therefore, responsiveness is rated as positive if hypotheses about the relationship between change in the instrument and corresponding changes in reference measurements were specified in advance. A positive rating is also given if the instrument is able to distinguish clinically important change from measurement error. Therefore, responsiveness must be tested by relating the SDC to the MIC, as described under Reliability.

Interpretability: Interpretability is defined as the degree to which (change) scores on an instrument can be interpreted. Mean scores and standard deviations should be reported for at least four relevant (sub) groups of patients. In addition, the authors must provide information about what (difference in) score would be clinically meaningful, and no floor or ceiling effects must be present. Floor and ceiling effects were considered to be present if more than 15% of the respondents achieved the highest or lowest possible score. If all the above mentioned requirements are met, interpretability is rated as positive.

Scoring of the measurement properties: For each of the above mentioned measurement properties the following rating options were used: 0 = not done, - = low quality, ? = indeterminate and + = high quality. Validity, reliability and responsiveness depend on the setting and the population in which they are assessed. Therefore, descriptions of the characteristics of the study population, measurements, setting and data analysis of every individual clinimetric study were rated. If a description was lacking or methodological weaknesses were found, the clinimetric property was rated as indeterminate.

Results

Selection of studies

The search strategy yielded a total of 2015 hits (Figure 1). The titles and abstracts were screened,

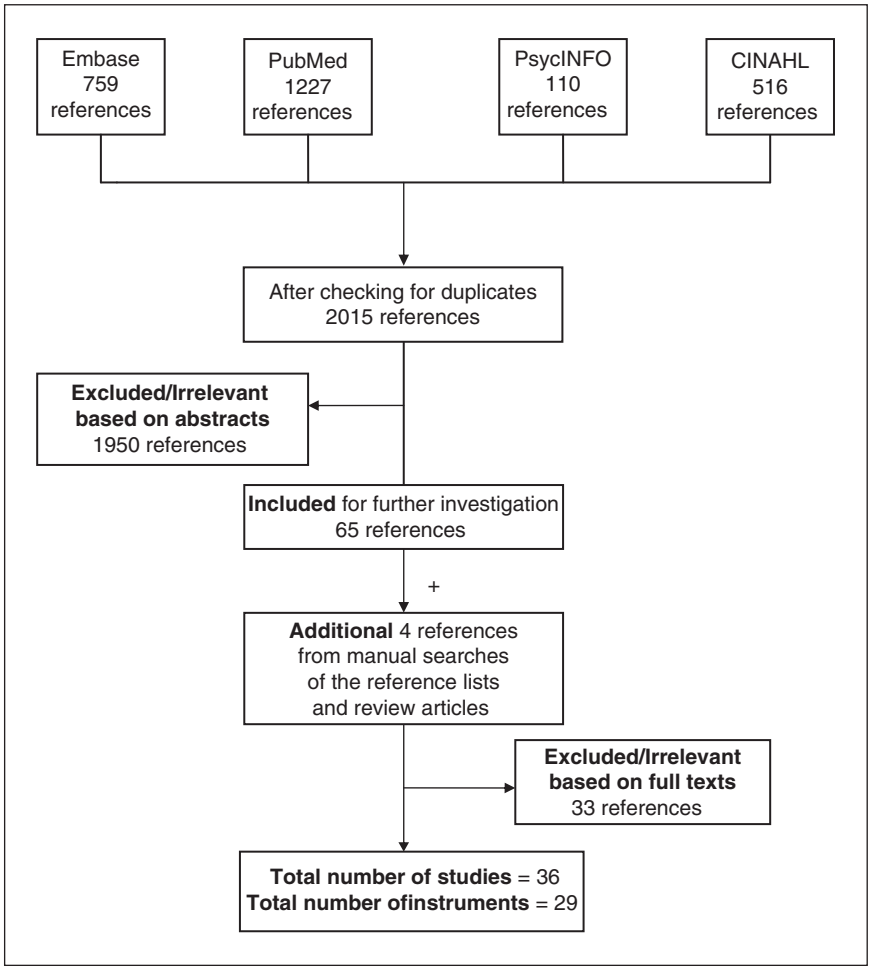

Figure 1. Results of the search strategy.

excluding 1950 references as irrelevant according to the inclusion and exclusion criteria described in the Methods section. The main search was supplemented by manual searches of the reference lists of the retrieved articles, which yielded four additional articles. Of the 69 full-text articles we studied, 36 met the inclusion criteria. Most of the excluded studies concerned quality-of-life instruments, but the evaluation of the measurement properties was not described.[20–28] Other studies were excluded because of an irrelevant study population, for example a curative patient population,[29–35] or because the aim of the study was not to develop or validate an instrument but, for example, to compare different questionnaires,[36–44] or because the instrument that was validated was not available in English or Dutch.[45–48] Another reason for exclusion was that the instrument was intended to measure the quality of the care or satisfaction with the care.[49–52] Finally, a total of 36 studies concerning 29 questionnaires was included in this review.

Instrument characteristics

Table 1 presents a description of the 29 instruments (full names are given in Appendix 3). More than half of the questionnaires were specifically developed for palliative care patients in general, but several questionnaires were designed for cancer patients, and two for hospice patients. The PNPC had the most items ($n = 138$), followed by the NA-ACP ($n = 132$), while the MQOL-CSF ($n = 8$) and the PAQ ($n = 4/9$) had the least items. The Emanuel and Emanuel medical directive could take two or three hours to complete, whereas the PDI, the CMSAS and the ESAS all take about two to five minutes to complete. Most of the instruments are self-report questionnaires, designed to be completed by the patient. The POS has two almost identical versions, a patient version and a staff version. Five other questionnaires could be completed by either the patient or a proxy. The Emanuel and Emanuel medical directive, the MRDI, the QODD and the QUAL-E are interview-based questionnaires. The SEIQoL[53] is not included in the tables because it differs from the other instruments with regard to the mode of administration (semi-structured interview) and the nature of the generated data (individual, patient-generated scores and dimensions). Therefore, the categories that apply to all other instruments presented in the tables do not apply to the SEIQoL.

Measurement properties

Table 2 presents the published clinimetric data concerning the identified questionnaires. The ratings of the measurement properties that were assigned to the instruments are shown in Table 3. None of the instruments included in our review had been adequately tested for all measurement properties on the rating list.

The MQOL had the best clinimetric quality rating, followed by the QUAL-E and the QODD. All these questionnaires have good content validity, construct validity and internal consistency, but only the MQOL has good reliability. Information on responsiveness, absolute measurement error and interpretability was lacking or insufficient for the MQOL, the QUAL-E and the QODD.

Discussion

Our review identified 29 questionnaires to assess the quality of life of palliative care patients, of which 7 were revised versions of the original instruments. The characteristics and the clinimetric quality of the instruments varied substantially. None of the instruments achieved satisfactory ratings for all categories. Overall, the MQOL received the best ratings for its measurement properties, followed by the QUAL-E and the QODD. These questionnaires are all designed to assess the quality of life of palliative care patients in general, but only the QODD is designed to be completed by family members or health care workers.

Because many measurement properties were not (adequately) tested for a large number of instruments, we describe the shortcomings of the testing below. In order to achieve adequate *content validity*, the involvement of the target population in the item selection is crucial, because patients are the experts on their own quality of life. The selection of items was inadequately performed for seven of the instruments, mainly because the patients were not involved in the process. Furthermore, 18 questionnaires fulfilled the requirements with regard to content validity.

Studies evaluating *construct validity* were available for all but four instruments. In all articles except one, construct validity was assessed by correlating the instrument to (sub scales of) other quality-of-life measures, performance scores or symptom distress scores. Nevertheless, 10 instruments scored 'doubtful' for construct validity because no hypotheses were formulated, and four other instruments scored doubtful because there was no information about the expected direction or magnitude of the correlation. Furthermore, when reviewing the articles, it is impossible to check whether hypotheses were formulated before the data analysis was performed.

When developing a questionnaire, the theoretical dimensional structure should be tested with factor-analysis, but this had not been done for six questionnaires included in this study. Another reason for a

Table 1. Description of the instruments

Instrument	Target population	Domains[a]	Number of items	Number of response options	Scoring algorithm	Recall period[b]	Completion time	Self-report/ proxy/ interview	Full copy of instrument available
BHI[55]	Hospice patients	Symptoms; QoL[a]	17	11	–	No/few days/week	9 min.	Self-report[c]	Yes
CAMPAS-R[56]	All kinds of palliative care patients	Presence and interference of physical and psychological symptoms	2 × 10	VAS	Sub scale	Week	–	Self-report	Yes
DS[57]	Cancer patients	Loss of meaning; dysphoria; disheartenment, helplessness; sense of failure[a]	24	5	Sub scale and total	2 weeks	–	Self-report	Yes
EFAT[58,59]	Cancer patients	Symptoms and functions	11	4	Total	No	–	Proxy	Yes
Emanuel and Emanuel Medical Directive[60]	Severely ill patients	*Patient-generated*	48	–	Not applicable	Not applicable	2–3 hours	Structured interview	No
EORTC QLQ-OES18[61]	Oesophageal cancer patients	Dysphagia; eating restrictions; pain; reflux; 6 single items	18	4	Sub scale	1 week	15 min. (incl. completion of EORTC QLQ-C30)	Self-report	Yes
EORTC QLQ-STO22[62]	Patients with adenoma carcinoma of the stomach	Dysphagia; eating restrictions; pain; reflux; anxiety; 3 single items	22	4	Sub scale	1 week	15 min (incl. completion of EORTC QLQ-C30)	Self-report	Yes
ESAS[63]	Palliative care patients	Symptoms	10	VAS 10	Total	No	5 min.	Self-report	Yes
FACIT-Pal[64]	Patients with life limiting illness	Symptoms; family and friend relationships; life closure issues; decision making and communication abilities	19	5	Total	7 days	–	Self-report	Yes
HQLI[65,66]	Hospice patients	Psychophysiological; functioning; social/spiritual[a]	28	11	Sub scale and total	–	10–15 min	Self-report	No
LCS[67]	Terminally ill patients	Self-reconciled; Self-restructuring[a]	20	5	Sub scale and total	–	–	Self-report	Yes
LEQ[68]	People with incurable cancer	Freedom; appreciation of life; contentment, resentment; social integration[a]	44	7	Sub scale	–	–	Self-report	No
MQLS[69]	Palliative care patients	Physical; cognition; social; energy; role; rest; function; emotion	32	7	–	24 hours	3–30 min	Self-report[c]	Yes
MQOL[70,71]	People with life threatening illness	Psychological symptoms; existential well-being; support; physical symptoms[a]	16	11	Sub scale and total	2 days	10–30 min	Self-report	Yes
MQOL-CSF[72]	Terminally ill patients	Global QoL; physical symptoms; psychological; existential	8	11	Sub scale and total	2/7 days	3.26 min	Self-report	No
MRDI[73]	Terminally ill patients	Physiological; psychological; sociological; spiritual	28	VAS	Total score	–	–	Structured interview[c]	No
MSAS[74,75]	Cancer patients	Global distress; physical; psychological[a]	32	4/5	Sub scale and total	1 week	20–60 min	Self-report/ proxy	Yes

CMSAS[76]	Cancer patients	Physical and psychological symptoms	14	4/5	Sub scale and total	7 days	2–4 min	Self-report	Yes
MSAS-GDI[77]	Cancer patients	Psychological symptoms; physical symptoms	11	4	Total	Last week of life	–	Proxy	Yes
MVQOLI-R[78,79]	Terminally ill patients	Symptoms; function; interpersonal; well-being; transcendence	25	5	Sub scale and total	No	–	Self-report[c]	Yes
NA-ACP[80]	Advanced cancer patients	Medical communication/information; psychological/emotional; daily living; financial; symptom; spiritual; social	132	5	Sub scale	4 months	76 min	Self-report	No
PAQ[81]	Palliative cancer patients	Autonomy	4/9	3	Total	–	–	Self-report	No
PDI[82]	Patients nearing the end of life	Symptom distress; existential distress; dependency; peace of mind; social support[a]	25	5	–	Few days	2 min (max. 10–15 min)	Self-report	Yes
PNPC[83]	Palliative care patients	ADL&IADL; physical symptoms; role activities; financial; social; psychological; spiritual; autonomy; informational needs; quality-of-care aspects	138	3	Sub scale	No	–	Self-report	Yes
PNPC-sv[84]	Palliative care patients	ADL&IADL; physical symptoms; financial; social; psychological; spiritual; autonomy; informational needs	33	3	Sub scale	No	–	Self-report	Yes
POS[85]	Advanced cancer patients and staff	Physical; psychological; spiritual	Patient: 10; staff: 12	3/4/5	–	3 days	Patients 6.9 min; staff 5.7 min	Self-report/proxy	Yes
QODD[86]	Family members of terminallyill patients/ health care workers	Symptoms and personal care; preparation for death, moment of death; family; treatment preferences; whole person concerns	31	11	Total	Retrospective	–	Structured proxy interview	Yes
QUAL-E[87]	Seriously ill patients	Life completion; relationship with health care provider; preparation[a]	26	5	Sub scale and total	1 week/ 1 month	–	Structured interview	Yes
SNI[88]	Patients near the end of life	Outlook; inspiration; spiritual activities; religion; community[a]	17	5	Sub scale and total	–	–	Self-report	Yes

[a]Domains are the same as factors determined by factor analysis.
[b]When several recall periods are reported, the recall period differs over items.
[c]Administration by proxy possible.
–: no information available.

Table 2. Identified instruments with published clinimetric data

Measure (reference)	Study population	Validity	Internal consistency	Reliability	Responsiveness
BHI[55]	Hospice patients	Overall symptom scores correlates with quality of life scores for patients ($r = 0.71$, $p < 0.001$)	Cronbach's α of the 2 subscales: 0.88; 0.94	Tested by paired sample *t*-test: 0.58–0.63 (time interval: $\pm$ 1 wk)	-
CAMPAS-R[56]	Palliative home care patient (estimated to be in their last year of life)	Pain severity and interference scores on CAMPAS-R correlated significantly with corresponding scores on BPI and EORTC ($r = 0.82$–0.88); All other physical severity and interference symptom scores on CAMPAS-R correlated also significantly with the EORTC ($r = 0.31$–0.91); Patient anxiety and depression strongly correlated with corresponding scores on HADS and EORTC ($r = 0.55$–0.66 resp. $r = -0.67$–0.77)	Cronbach's α severity: 0.77; Cronbach's α interference: 0.80	-	Predefined hypotheses were tested: CAMPAS-R could differentiate between patients who were near vs. less near to death (cut-off point: 60 days)
DS[57]	In patients with cancer	Total scores and subscales significantly correlates with the McGill QOL, BDI, PHQ, Beck's measure of hopelessness, and SAHD to measure desire for hastened death; Demoralization differentiates from depression	Cronbach's α: 0.70–0.89	-	-
EFAT[58]	Cancer patients from a palliative care unit	Predefined hypotheses were tested: EFAT scores increase from admission to discharge: *confirmed;* EFAT scores are both higher at admission and discharge for patients remained until death or transferred to another institution than for patients discharged home: *confirmed*; Correlated with KPS ($r = -0.79$, $p = 0.001$), ECOG ($r = 0.85$, $p = 0.0001$) and global PS rating ($r = 0.90$, $p = 0.0001$); Correlations for 7/10 EFAT items to total score ranged from $r = 0.62$–0.95	-	-	-
EFAT-2[59]	Cancer patients from a palliative care unit	Predefined hypotheses were tested: Patients transferred to hospice have higher scores (poorer functional status) than patients discharged home: *confirmed* All correlations between items were ≥ 0.3 except the pain item which not correlated with any other item	Cronbach's α: 0.86	-	-
Emanuel and Emanuel Medical Directive[60]	In patients with progressive, chronic and life-threatening illness and a prognosis between 6 wks and 3 yrs (dialysis clinics; rehabilitation hospitals; long-term facilities)	Predefined hypotheses were tested: Hospice patients decided more frequently to forego curative or aggressive life-sustaining treatments: *confirmed* Consistently lower preference scores across situations for those with goal "comfort care" vs "prolong life" ($p < 0.0001$)	Cronbach's α across treatments by scenario: 0.80–0.85 Cronbach's α across scenarios by treatment: 0.86–0.90	ICC across treatments by scenario: 0.69–0.75 ICC across scenarios by treatment: 0.60–0.79 (time interval: 21 days)	Small effect size for change in preferences as a function of health status change (patients with worsened health status, want less intervention, whereas those improved health want more intervention)

EORTC QLQ-OES18[61]	Patients with oesophageal squamous cell or adenocarcinoma	Predefined hypotheses were tested: Correlation of >0.40 between items and its own scale: *confirmed;* A higher correlation between an item and its own scale than with another scale: *confirmed;* No correlation between EORTC QLQ-OES18 scales and generic aspects of QoL unless they address similar themes: *confirmed* EORTC QLQ-OES18 could differentiate between curative and palliative patients in several functional and symptoms scales; Could differentiate between surgery patients and chemoradiation patients	Cronbach's α: 0.61–0.75	-	EORTC QLQ-OES18 could demonstrate treatment-induced changes over time
EORTC QLQ-STO22[62]	Patients with gastric cancer	Predefined hypotheses were tested: Correlation of >0.40 between items and its own scale: *confirmed;* A higher correlation between an item and its own scale than with another scale: *confirmed* EORTC QLQ-STO22 could differentiate in several scales and items between palliative and curative patients and between patients with Karnofsky scores <80 and >80	Cronbach's α: 0.72–0.80	-	The reflux scale demonstrated sensitivity to changes in weight loss over time; Dysphagia, pain, reflux and eating were sensitive to changes in observer-rated dysphagia scores; Could demonstrate treatment-induced changes over time
ESAS[63]	In and out patients with cancer	ESAS distress score correlated most closely with physical symptom subscales in the FACT and MSAS and with KPS	Cronbach's α of the overall ESAS: 0.79	Spearman's test-retest correlations: 0.86 ($p<0.0001$); 0.45 ($p<0.05$) (time interval: resp. 2 days; 1 wk)	-
FACIT-Pal[64]	Cancer patients	Predefined hypotheses were tested: Patients died within 3 months and patients lived ≥ 1 year differ significantly on the FACT-G subscales and FACIT-Pal: *confirmed;* Patients with KPS score ≥ 80 have higher QoL than patients with KPS ≥ 70: *confirmed*; Overall QoL and physical, functional and palliative subscales correlates with symptom intensity and depression: *confirmed*	Cronbach's α: 0.75–0.85	-	-
HQLI[65]	Hospice home care patients with cancer	Predefined hypotheses were tested: A weak significant correlation with ECOG-PSR: *confirmed;* HQLI could differentiate between patients and non patients: *confirmed*	Cronbach's α: 0.82–0.85	-	-
HQLI[66]	Hospice home care patients with end-stage cardiac disease	Predefined hypotheses was tested: A weak-to-moderate negative correlation with the MSAS global symptom distress score: *confirmed* Significant correlation between HQLI psychopsychological well-being scale and MSAS ($r=-0.45$, $p=0.012$)	Cronbach's α of the overall HQLI: 0.78	-	-

(continued)

Table 2. Continued

Measure (reference)	Study population	Validity	Internal consistency	Reliability	Responsiveness
LCS[67]	Hospice patients diagnosed with cancer; AIDS; ALS; end-stage system diseases	Correlated with the total score as well as with the subscales of the Affect Balance Scale	Cronbach's α for 2 subscales: 0.80; 0.82	-	-
LEQ[68]	Cancer patients from outpatient clinics; hospital wards	Correlated with RSC: 0.01–0.62 (<0.01 only for resentment, freedom and social integration); Correlated with MacAdam and Smith: 0.02–0.62 (<0.01 only for resentment, freedom and social integration)	Cronbach's α: 0.70–0.85	Product-moment correlations: 0.77–0.92 (time interval: 48–72 h)	-
MQLS[69]	In and out, and community-based patients	Predefined hypotheses were tested: Correlation with the Spitzer Quality of Life Index: *Confirmed;* Patients able to rate the MQLS themselves rated QoL higher than those who needed assistance: *Confirmed;* Correlation with length of time until death and not with age/sex of the patient: *Confirmed;* Non-symptom-related aspects of QoL are lower among those within 3 weeks of death than among survivors, while symptom ratings would be more similar: *Confirmed*	Cronbach's α for patients: 0.09–0.69 (all nonphysical items combined: 0.79; overall scale: 0.8) Cronbach's α for family: 0.21–0.63 (all nonphysical items combined: 0.84; overall scale: 0.87) Cronbach's α for staff: 0.50–0.78 (all nonphysical items combined: 0.88; overall scale: 0.89)	$r = 0.63$, CI = 0.45–0.77 (time interval: 1 wk)	Patients were asked if a change had occurred in QoL since the first rating, 1 wk before: Sensitive to changes in patient's QoL ($F = 5.26$, $df = 2.53$, $p = 0.01$)
MQOL[70]	Patients from palliative care inpatient units (hospital)	Predefined hypotheses were tested: Global measure of QoL (single item) correlates higher MQOL total score than with Spitzer QLI: *confirmed*; MQOL physical measures correlates most highly with Spitzer QLI activity, daily living, health items: *not confirmed*; MQOL psychological subscale correlates most highly with Spitzer QLI health, outlook items: confirmed; MQOL existential subscale correlates most highly with any Spitzer QLI outlook item: *confirmed*; MQOL support subscale correlates most highly with the Spitzer QLI support item: *confirmed*	Cronbach's α: >0.70 except physical subscale (0.62)	-	-
MQOL[71]	Cancer patients from oncology day centre, receiving chemotherapy; palliative home care service	-	-	ICC: 0.62–0.85 (time interval: 2 days)	Patients ranked their days as good, average, or bad: Pearson correlations: MQOL-change score: 0.56 (existential) - 0.66 (total), except for support scale (0.13); Effect sizes: largest for differences between good/bad days; moderate between bad/average days; average/good days, except for physical resp. support scale

MQOL-CSF[72]	Patients from hospice centre (reliability test); hospital inpatient wards (validity test)	Predefined hypothesis were tested: High correlation with MQOL-CSF items: *confirmed* (0.48–0.73); Strong correlations (>0.40) between MQOL-SCF items and their own domains: *confirmed* (4/7); Items correlates strongest with their own domain: (6/7);Weak correlation with SF-36 general health question: *confirmed* (0.18–0.40); Low haemo globin levels are associated with lower MQOL-CSF scores: *confirmed*	Cronbach's α: 0.64–0.81, except existential domain (0.46)	Spearman's test-retest correlations: 0.5–0.86 (time interval: 1 wk)	-
MRDI[73]	Hospice patients	Correlated with overall readiness for death question ($r = 0.45$, $p < 0.01$); Correlation between the scores of patients and primary caregivers ($r = 0.35$, $p < 0.05$); between patients and primary hospice nurses ($r = 0.53$, $p < 0.01$); Significant mean difference between terminally-ill and non-terminally cardiac-impaired patients ($t = 2.76$, $p < 0.01$)	Cronbach's α of the over all MRDI: 0.59 (holistic measure)	Test-retest reliability: $r(12) = 0.22$, $p = 0.22$ (low because $n = 70$) (time interval: 7–14 days)	-
MSAS[74]	Advanced cancer and AIDS patients and their family caregivers actively involved in their care	Moderate-to-strong correlations between the GDI and physical and psychological symptoms (range, 0.58–0.81, $p = 0.002$) as well as moderate correlations between physical and psychological symptoms (range, 0.47–0.60, $p = 0.001$) for all patients and caregivers groups	Cronbach's α AIDS patients; caregivers: 0.78–0.87; 0.86–0.91 Cronbach's α Cancer patients; caregivers: 0.78–0.83; 0.81–0.86	-	-
MSAS (FC)[75]	Family caregivers of cancer patients	Inter-item correlation was $r = 0.45$ on psychological scale; $r = 0.30$ on physical scale; $r = 0.35$ on MSAS-GDI; Item-scale correlation was $r = 0.60$ on psychological scale; $r = 0.50$ on physical scale; $r = 0.54$ on MSAS-GDI; Correlations between patients' ratings (regarded as gold standard) and family caregivers' ratings: Kappa = 0.22–0.70 physical subscale; Kappa = 0.16–0.48 psychological subscale; ICC = 0.68 physical subscale; ICC = 0.32 psychological subscale; ICC = 0.82 MSAS-GDI	Cronbach's α: 0.82–0.84	-	-
CMSAS[76]	In and out cancer patients	Subscales correlated significantly with FACT-G subscales (−0.61–0.76), KPS (−0.31–0.64), and MSAS-SF (0.89–0.93)	Cronbach's α: 0.72–0.85	-	-
MSAS-GDI[77]	Family members of cancer patients died in hospital, involved in the decedent's care and decision making in the final month of life	Item-total correlation: $r = 0.49$; Average inter-item correlations: $r = 0.30$; Higher MSAS-GDI scores correlates with discomfort on a single-item indicator ($r = 0.26$, $p = 0.02$); and with dissatisfaction with help from other family members on a single-item indicator ($r = -0.21$, $p = 0.05$)	Cronbach's α of the overall MSAS-GDI: 0.82	-	-
MVQOLI[78]	Hospice patients	Correlated with MQOLS-CA2, Pearson correlation coefficient: 0.63; MVQOLI total score correlated with a global QoL rating (0.43); Low correlation with KPS (0.19)	Cronbach's α of the overall MVQOLI: 0.77	-	-

(continued)

Table 2. Continued

Measure (reference)	Study population	Validity	Internal consistency	Reliability	Responsiveness
MVQOLI-R[79]	Patients from dialysis clinics; hospices; long-term care facilities	Predefined hypotheses were tested: Moderately high correlations with measures of psychological well-being, global symptom-related distress and global QoL but weak with mood and total symptom burden; *not confirmed*	Cronbach's α: 0.23–0.70 Factor analysis showed a heterogeneous structure, items were not effect indicators, but causal indicators	ICC: 0.59–0.77 (time interval: 3–5 days)	Predefined hypotheses were tested: Lower levels of MVQOLI-R are associated with worse global symptom related distress, independent of changes in mood: *confirmed*
NA-ACP[80]	Advanced, incurable cancer patients, not currently receiving formal palliative care	-	Cronbach's α: 0.79–0.98	Test-retest reliability: ICC: 0.67–0.93 Agreement: K: 0.18–0.83 (K = 0.4–0.2 for 28 of 132 items) (time interval: 1 wk)	-
PAQ[81]	Cancer patients	Predefined hypotheses were tested: Moderate to high correlations with EORTC QLQ-C30 and COOP-WONCA: *confirmed*; Scoring in the highest quartile of the PNPC is associated with more autonomy problems compared to scoring in the lowest quartile; *confirmed*	Cronbach's α of the 9-item version: 0.86 Cronbach's α of the 4-item version: 0.71	-	-
PDI[82]	In patients receiving palliative care	Predefined hypotheses were tested: The symptom distress scale correlates significantly with all ESAS items except for activity; and with will to live, the General Well-Being scale, the Beck Depression Inventory and the single-item measure of suffering: *confirmed*; The existential distress scale correlates with measures of depression, suffering, well-being, quality of life, satisfaction with quality of life, but not with will to live and loss of dignity: *confirmed*; The dependency scale correlates with activity, ability to work, rating and satisfaction with quality of life and sense of dignity: *confirmed*; The peace of mind scale correlates only with anxiety, the Beck Depression Inventory and FACIT inner peace factor, but not with will to live, suffering, well-being, rating and satisfaction with quality of life, FACIT total, faith/spirituality and meaning and spirituality factors: *not confirmed;* The social support scale cor relates with availability of, and satisfaction with support: *confirmed*	Cronbach's α: 0.63–0.83	Test-retest reliability: $r = 0.85$ for the full PDI; $r = 0.37$–0.76 with individual variables	-
PNPC[83]	Cancer patients, living at home	Predefined hypotheses tested: Substantial correlation with related quality of life dimensions of EORTC QLQ-C30 and COOP-WONCA: *confirmed*	Cronbach's α: 0.67–0.89 (problem aspect)	-	-

PNPC-sv[84]	Cancer patients, living at home	Predefined hypotheses tested: Correlation ($\rho > 0.80$) with corresponding dimensions in the original PNPC: *confirmed* (ρ: 0.83–0.98); Correlation ($\rho > 0.40$) with corresponding HRQoL dimensions of the EORTC QLQ-C30 and COOP-WONCA: *confirmed*	Cronbach's α: 0.61–0.86 (problem aspect)	-	-
POS[85]	Patients from centres providing palliative care, including inpatient, outpatient, day, home and primary care	Predefined hypotheses tested: Patient version correlates with EORTC QLQ-C30 (physical symptoms, all non QoL problems, QoL): *confirmed* ($\rho = 0.43$–0.53); Staff version correlates with STAS (physical symptoms, all non QoL problems, QoL): *confirmed* ($\rho = 0.51$–0.80)	Cronbach's α patient version: 0.65 Cronbach's α staff version: 0.70	Kappa: 0.08-0.62 (time interval was dependent on nature of service providing care)	Change over time was not statistically significant
QODD[86]	Dying patients	Predefined hypothesis tested: Higher QODD-scores correlates with death at home, death in location patient desired, lower symptom burden, better ratings for symptom treatment, adherence to patients preferences for end-of-life care, compliance with treatment preferences, family satisfaction regarding communication with health care team, availability of health care team member at night/weekends: *confirmed*	Cronbach's α for overall QODD 0.89 Factor analysis did not support subscale construction		
QUAL-E[87]	Patients with cancer, congestive heart failure, end stage renal disease, chronic obstructive pulmonary disease	Predefined hypotheses were tested: Strong correlation (> 0.60) between FACIT-SP spiritual well-being subscale and QUAL-E: *confirmed;* Moderate correlation (0.4–0.6) between QUAL-E completion and preparation subscale and FACIT-SP: *confirmed;* Moderate associations between QUAL-E and MVQOL's similar domains: *confirmed;* Moderate correlation between first two PDM items and QUAL-E health care subscale: *confirmed;* Correlations among QUAL-E subscales and no relationships among unrelated subscales and weak to moderate correlations among related scales or those with conceptual overlap: *confirmed*	Cronbach's α: 0.68–0.87	Test-retest reliability: r: 0.61–0.74 except for symptoms r: 0.23	
SNP[88]	Patients from outpatient hospices and one inpatient hospice facility	Predefined hypotheses was tested: The number of unmet needs correlates with life satisfaction as measured by a Cantril ladder: *Confirmed:* −0.17 Item-to-total correlations: 0.33–0.70	Cronbach's α: 0.62–0.78	-	-

–: no data published.

Table 3. Rating of measurement properties of the instruments

Instrument	Content validity	Construct validity	Internal consistency	Reliability	Absolute measurement error	Responsiveness	Interpretability
BHI[55]	?	0	+	?	0	0	0
CAMPAS-R[56]	+	?	?	0	0	?	0
DS[57]	?	?	?	0	0	0	0
EFAT[58]	+	+	0	0	0	0	0
EFAT-2[59]	+	?	+	0	0	0	0
Emanuel and Emanuel Medical Directive[60]	?	?	?	-	0	?	?
EORTC QLQ-OES18[61]	+	?	-	0	0	?	?
EORTC QLQ-STO22[62]	+	?	+	0	0	?	?
ESAS[63]	?	?	?	?	0	0	0
FACIT-Pal[64]	?	+	?	0	0	0	0
HQLI[65]	?	+	+	0	0	0	0
HQLI (in end stage cardiac disease patients)[66]	?	+	?	0	0	0	0
LCS[67]	?	?	?	0	0	0	0
LEQ[68]	?	?	?	?	0	0	?
MQLS[69]	+	+	?	?	0	?	0
MQOL[70,71]	+	+	+	+	0	?	0
MQOL-CSF[72]	?	+	?	?	0	0	0
MRDI[73]	?	?	?	?	0	0	0
MSAS[74]	+	0	?	0	0	0	0
MSAS (FC)[75]	+	?	?	0	0	0	0
CMSAS[76]	+	?	?	0	0	0	0
MSAS-GDI[77]	+	0	+	0	0	0	0
MVQOLI[78]	+	+	0	0	0	0	0
MVQOLI-R[79]	+	?	-	?	0	?	0
NA-ACP[80]	?	0	?	+	?	0	0
PAQ[81]	?	+	?	0	0	0	0
PDI[82]	?	+	+	?	0	0	0
PNPC[83]	+	+	?	0	0	0	0
PNPC-sv[84]	+	+	?	0	0	0	0
POS[85]	?	?	?	-	0	?	0
QODD[86]	+	+	+	0	0	0	?
QUAL-E[87]	+	+	+	?	0	0	0
SNI[88]	+	?	?	0	0	0	0

Method or result was rated as: +: high quality, ?: indeterminate, -: low quality, 0: no data available.

doubtful rating for *internal consistency* was an inadequate study size. Moreover, Cronbach's alpha is positively influenced by the number of items in a sub scale, irrespective of the average correlation among items. Five out of nine questionnaires which were rated positive for internal consistency in this study contained more than 22 items. Furthermore, for almost all questionnaires it was not clear whether the items were based on a reflective model or a causal model.

For 12 instruments a test–retest study was performed, but only two questionnaires met our criteria for good *reliability*. Several authors calculated a correlation coefficient, but this measure is inadequate because systematic differences are not taken into account. Moreover, because terminally ill patients are rarely stable, it is complicated to determine an adequate time-interval between measurements. A short time-interval (>1 week) often causes recall bias, but

palliative care patients may change with regard to the construct to be measured if the time-interval is more than one week.

All the instruments identified in this review were developed as an evaluative outcome measure. However, the *responsiveness* of quality-of-life questionnaires is seldom tested. None of the instruments had adequate responsiveness, but this is probably due to the strictness of the criteria for testing responsiveness. Moreover, the MIC and the SDC are relatively new concepts that have received much attention recently. However, a considerable number of quality-of-life instruments were developed and validated before there was consensus on the criteria for testing responsiveness. The same applies to *absolute measurement error*, which was not calculated for one of the identified questionnaires.

None of the developers of the questionnaires included in this review paid sufficient attention to the *interpretability* of the outcome scores, which is not remarkable given the strict criteria for interpretability. It is difficult to recruit sufficient terminally ill patients, let alone to recruit four relevant sub groups of patients.

We set high standards for the assessment of measurement properties and, accordingly, many measurement properties were not favourably evaluated. However, 'doubtful' or 'poor' ratings for the clinimetric characteristics of a questionnaire do not necessarily mean that the questionnaire is inadequate. A doubtful rating should be a motive for further testing and evaluating the measurement properties according to the criteria developed by Terwee et al.[13] Therefore, our intention is not to promote the development of new quality-of-life questionnaires for use in palliative care, but to support further testing of existing instruments with good content validity and to select one or a few which are most appropriate for clinical use and/or research purpose. In order to improve palliative care nationally and internationally, organizations for the promotion and development of palliative care, such as the European Association for Palliative Care (EAPC) or the International Association for Hospice & Palliative Care (IAHPC), should also support further testing of the existing quality-of-life instruments, which would also benefit all researchers working in this field. An important advantage of the use of one or a few well-developed and adequately tested questionnaires is the comparability of research results.

This study has a few limitations. First, many studies were identified by our review, but we cannot be sure that we did not miss any. However, the search strategy included a clinimetric search filter with a sensitivity of 90–97% to retrieve clinimetric articles, so it is unlikely that we missed any relevant articles.[54] Furthermore, we checked the references of the articles we included and we also consulted some experts to ensure we had not missed any instruments. Another limitation could be the restriction to the English and Dutch languages. However, because measurement properties are not automatically stable across different languages or cultures, an instrument should be tested in the target population and language, in accordance with the aim of study.

In conclusion, we presented a systematic review of 29 questionnaires which measured (at least one domain of) quality of life applicable in the palliative care setting. Information about practical aspects, such as the burden for the respondent, and the clinimetric quality of these instruments could help clinicians and researchers in their choice of measurement instrument. Apart from the clinimetric quality of the instrument, the purpose of the study also plays a role in the choice of an instrument. If the purpose of the measurement is evaluation, testing for responsiveness is important, and if the purpose of the study is discrimination, reliability testing is of significance. As a consequence, we cannot provide an explicit recommendation for the use of one specific instrument. Future research should focus on further testing of these measurement instruments.

References

1. Jocham HR, Dassen T, Widdershoven G, Halfens R. Quality of life in palliative care cancer patients: a literature review. *J Clin Nurs* 2006; 15: 1188–1195.
2. Stewart AL, Teno J, Patrick DL, Lynn J. The concept of quality of life of dying persons in the context of health care. *J Pain Symptom Manage* 1999; 17: 93–108.
3. Bruley DK. Beyond reliability and validity: analysis of selected quality-of-life instruments for use in palliative care. *J Palliat Med* 1999; 2: 299–309.
4. Hearn J, Higginson IJ. Outcome measures in palliative care for advanced cancer patients: a review. *J Public Health Med* 1997; 19: 193–199.
5. Jordhoy MS, Inger RG, Helbostad JL, Oldervoll L, Loge JH, Kaasa S. Assessing physical functioning: a systematic review of quality of life measures developed for use in palliative care. *Palliat Med* 2007; 21: 673–682.
6. Kaasa S, Loge JH. Quality-of-life assessment in palliative care. *Lancet Oncol* 2002; 3: 175–182.
7. Kirkova J, Davis MP, Walsh D, et al. Cancer symptom assessment instruments: a systematic review. *J Clin Oncol* 2006; 24: 1459–1473.
8. Massaro T, McMillan SC. Instruments for assessing quality of life in palliative care settings. *Int J Palliat Nurs* 2000; 6: 429–433.
9. McMillan SC. Quality-of-life assessment in palliative care. *Cancer Control* 1996; 3: 223–229.

10. Mularski RA, Dy SM, Shugarman LR, et al. A systematic review of measures of end-of-life care and its outcomes. *Health Serv Res* 2007; 42: 1848–1870.
11. Steinhauser KE, Christakis NA, Clipp EC, McNeilly M, McIntyre L, Tulsky JA. Factors considered important at the end of life by patients, family, physicians, and other care providers. *JAMA* 2000; 284: 2476–2482.
12. Steinhauser KE, Voils CI, Clipp EC, Bosworth HB, Christakis NA, Tulsky JA. ''Are you at peace?'': one item to probe spiritual concerns at the end of life. *Arch Intern Med* 2006; 166: 101–105.
13. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34–42.
14. Mokkink LB, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Med Res Methodol* 2006; 6: 2.
15. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther* 1996; 18: 979–992.
16. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993; 118: 622–629.
17. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985; 38: 27–36.
18. Fayers PM, Hand JH. Causal variables, indicator variables and measurement scales: an example from quality of life. *J R Stat Soc Ser A Stat Soc* 2002; 165: 233–261.
19. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res* 2003; 12: 349–362.
20. Arraras JI, Kuljanic-Vlasic K, Bjordal K, et al. EORTC QLQ-INFO26: a questionaire to assess information given to cancer patients a preliminary analysis in eight countries. *Psycho-Oncology* 2007; 16: 249–254.
21. Barresi MJ, Shadbolt B, Byrne D, Stuart-Harris R. The development of the Canberra symptom scorecard: a tool to monitor the physical symptoms of patients with advanced tumours. *BMC Cancer* 2003; 3: 32.
22. Groenvold M, Petersen MA, Aaronson NK, et al. The development of the EORTC QLQ-C15-PAL: a shortened questionnaire for cancer patients in palliative care. *Eur J Cancer* 2006; 42: 55–64.
23. Hoekstra J, Bindels PJ, van Duijn NP, Schade E. The symptom monitor. A diary for monitoring physical symptoms for cancer patients in palliative care: feasibility, reliability and compliance. *J Pain Symptom Manage* 2004; 27: 24–35.
24. Kearsley JH, Schonfeld C, Sheehan M. Quality-of-life assessment during palliative radiotherapy. *Australas Radiol* 1998; 42: 354–359.
25. Lehna C, Seck L, Churches S. Development of an outcome measure to document pain relief for home hospice patients: a collaboration between nursing education and practice. *Am J Hosp Palliat Care* 1998; 15: 343–351.
26. Mularski RA, Heine CE, Osborne ML, Ganzini L, Curtis JR. Quality of dying in the ICU: ratings by family members. *Chest* 2005; 128: 280–287.
27. Schaafsma J, Osoba D. The Karnofsky Performance Status Scale re-examined: a cross-validation with the EORTC-C30. *Qual Life Res* 1994; 3: 413–424.
28. Steel K, Ljunggren G, Topinkova E, et al. The RAI-PC: an assessment instrument for palliative care in all settings. *Am J Hosp Palliat Care* 2003; 20: 211–219.
29. Cella D. The Functional Assessment of Cancer Therapy-Lung and Lung Cancer Subscale assess quality of life and meaningful symptom improvement in lung cancer. *Semin Oncol* 2004; 31(3 Suppl 9): 11–15.
30. Cleeland CS, Mendoza TR, Wang XS, et al. Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory. *Cancer* 2000; 89: 1634–1646.
31. Gaunthier DM, Froman RD. Preferences for care near the end of life: scale development and validation. *Res Nurs Health* 2001; 24: 298–306.
32. Hollen PJ, Gralla RJ, Liepa AM, Symanowski JT, Rusthoven JJ. Adapting the Lung Cancer Symptom Scale (LCSS) to mesothelioma: using the LCSS-Meso conceptual model for validation. *Cancer* 2004; 101: 587–595.
33. King M, Jones L, Barnes K, et al. Measuring spiritual belief: development and standardization of a beliefs and values scale. *Psychol Med* 2006; 36: 417–425.
34. Kruse BG, Heinemann D, Moody L, Beckstead J, Conley CE. Psychometric properties of the Serenity Scale. *J Hosp Palliat Nurs* 2005; 7: 337–344.
35. Schirm V, Sheehan D, Zeller RA. Preferences for care near the end of life: instrument validation for clinical practice. *Crit Care Nurs Q* 2008; 31: 24–32.
36. Beadle GF, Yates PM, Najman JM, et al. Illusions in advanced cancer: the effect of belief systems and attitudes on quality of life. *Psycho-Oncology* 2004; 13: 26–36.
37. Bjorner JB, Petersen MA, Groenvold M, et al. Use of item response theory to develop a shortened version of the EORTC QLQ-C30 emotional functioning scale. *Qual Life Res* 2004; 13: 1683–1697.
38. Petersen MA, Groenvold M, Aaronson N, et al. Item response theory was used to shorten EORTC QLQ-C30 scales for use in palliative care. *J Clin Epidemiol* 2006; 59: 36–44.
39. Pratheepawanit N, Salek MS, Finlay IG. The applicability of quality-of-life assessment in palliative care: comparing two quality-of-life measures. *Palliat Med* 1999; 13: 325–334.
40. Sprangers MA, Cull A, Groenvold M, Bjordal K, Blazeby J, Aaronson NK. The European Organization for Research and Treatment of Cancer approach to developing questionnaire modules: an update and overview. EORTC Quality of Life Study Group. *Qual Life Res* 1998; 7: 291–300.
41. Stockler MR, Osoba D, Goodwin P, Corey P, Tannock IF. Responsiveness to change in health-related quality of life in a randomized clinical trial: a comparison of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) with analogous scales from the EORTC QLQ-C30 and a trial specific module. European Organization for Research and Treatment of Cancer. *J Clin Epidemiol* 1998; 51: 137–145.

42. Higginson IJ, McCarthy M. A comparison of two measures of quality of life: their sensitivity and validity for patients with advanced cancer. *Palliat Med* 1994; 8: 282–290.
43. Higginson IJ, Donaldson N. Relationship between three palliative care outcome scales. *Health Qual Life Outcomes* 2004; 2: 68.
44. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *J Clin Epidemiol* 2007; 60: 448–455.
45. Llobera J, Esteva M, Benito E, et al. Quality of life for oncology patients during the terminal period. Validation of the HRCA-QL index. *Support Care Cancer* 2003; 11: 294–303.
46. Maunsell E, Allard P, Dorval M, Labbe J. A brief pain diary for ambulatory patients with advanced cancer: acceptability and validity. *Cancer* 2000; 88: 2387–2397.
47. Miyashita M, Morita T, Sato K, Hirai K, Shima Y, Uchitomi Y. Good Death Inventory: a measure for evaluating good death from the bereaved family member's perspective. *J Pain Symptom Manage* 2008; 35: 486–498.
48. Mystakidou K, Tsilika E, Kouloulias V, et al. The ''Palliative Care Quality of Life Instrument (PQLI)'' in terminal cancer patients. *Health Qual Life Outcomes* 2004; 2: 8.
49. Carson MG, Fitch MI, Vachon ML. Measuring patient outcomes in palliative care: a reliability and validity study of the Support Team Assessment Schedule. *Palliat Med* 2000; 14: 25–36.
50. Casarett D, Pickard A, Bailey FA, et al. A nationwide VA palliative care quality measure: the family assessment of treatment at the end of life. *J Palliat Med* 2008; 11: 68–75.
51. Engelberg R, Downey L, Curtis JR. Psychometric characteristics of a quality of communication questionnaire assessing communication about end-of-life care. *J Palliat Med* 2006; 9: 1086–1098.
52. Higginson IJ, 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.
53. Waldron D, O'Boyle CA, Kearney M, Moriarty M, Carney D. Quality-of-life measurement in advanced cancer: assessing the individual. *J Clin Oncol* 1999; 17: 3603–3611.
54. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009; 18(8): 1115–1123.
55. Guo H, Fine PG, Mendoza TR, Cleeland CS. A preliminary study of the utility of the brief hospice inventory. *J Pain Symptom Manage* 2001; 22: 637–648.
56. Ewing G, Todd C, Rogers M, Barclay S, McCabe J, Martin A. Validation of a symptom measure suitable for use among palliative care patients in the community: CAMPAS-R. *J Pain Symptom Manage* 2004; 27: 287–299.
57. Kissane DW, Wein S, Love A, Lee XQ, Kee PL, Clarke DM. The Demoralization Scale: a report of its development and preliminary validation. *J Palliat Care* 2004; 20: 269–276.
58. Kaasa T, Loomis J, Gillis K, Bruera E, Hanson J. The Edmonton Functional Assessment Tool: preliminary development and evaluation for use in palliative care. *J Pain Symptom Manage* 1997; 13: 10–19.
59. Kaasa T, Wessel J. The Edmonton Functional Assessment Tool: further development and validation for use in palliative care. *J Palliat Care* 2001; 17: 5–11.
60. Schwartz CE, Merriman MP, Reed GW, Hammes BJ. Measuring patient treatment preferences in end-of-life care research: applications for advance care planning interventions and response shift research. *J Palliat Med* 2004; 7: 233–245.
61. Blazeby JM, Conroy T, Hammerlid E, et al. Clinical and psychometric validation of an EORTC questionnaire module, the EORTC QLQ-OES18, to assess quality of life in patients with oesophageal cancer. *Eur J Cancer* 2003; 39: 1384–1394.
62. Blazeby JM, Conroy T, Bottomley A, et al. Clinical and psychometric validation of a questionnaire module, the EORTC QLQ-STO 22, to assess quality of life in patients with gastric cancer. *Eur J Cancer* 2004; 40: 2260–2268.
63. Chang VT, Hwang SS, Feuerman M. Validation of the Edmonton Symptom Assessment Scale. *Cancer* 2000; 88: 2164–2171.
64. Lyons KD, Bakitas M, Hegel MT, Hanscom B, Hull J, Ahles TA. Reliability and validity of the Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-Pal) Scale. *J Pain Symptom Manage* 2009; 37: 23–32.
65. McMillan SC, Weitzner M. Quality of life in cancer patients: use of a revised Hospice Index. *Cancer Pract* 1998; 6: 282–288.
66. McMillan SC, Dunbar SB, Zhang W. Validation of the Hospice Quality-of-Life Index and the Constipation Assessment Scale in end-stage cardiac disease patients in hospice care. *J Hosp Palliat Nurs* 2008; 10: 106–117.
67. Dobratz MC. The life closure scale: additional psychometric testing of a tool to measure psychological adaptation in death and dying. *Res Nurs Health* 2004; 27: 52–62.
68. Salmon P, Manzi F, Valori RM. Measuring the meaning of life for patients with incurable cancer: the life evaluation questionnaire (LEQ). *Eur J Cancer* 1996; 32A: 755–760.
69. Sterkenberg CA, King B, Woodward CA. A reliability and validity study of the McMaster Quality of Life Scale (MQLS) for a palliative population. *J Palliat Care* 1996; 12: 18–25.
70. Cohen SR, Mount BM, Bruera E, Provost M, Rowe J, Tong K. Validity of the McGill Quality of Life Questionnaire in the palliative care setting: a multicentre Canadian study demonstrating the importance of the existential domain. *Palliat Med* 1997; 11: 3–20.
71. Cohen SR, Mount BM. Living with cancer: 'Good' days and 'Bad' days – what produces them? Can the McGill Quality of Life Questionnaire distinguish between them? *Cancer* 2000; 89: 1854–1865.

72. Lua PL, Salek S, Finlay I, Lloyd-Richards C. The feasibility, reliability and validity of the McGill Quality of Life Questionnaire-Cardiff Short Form (MQOL-CSF) in palliative care population. *Qual Life Res* 2005; 14: 1669–1681.
73. McCanse RP. The McCanse Readiness for Death Instrument (MRDI): a reliable and valid measure for hospice care. *Hosp J* 1995; 10: 15–26.
74. Sherman DW, Ye XY, Beyer MC, Parkas V, Calabrese M, Gatto M. Symptom assessment of patients with advanced cancer and AIDS and their family caregivers: the results of a quality-of-life pilot study. *Am J Hosp Palliat Care* 2007; 24: 350–365.
75. Lobchuk MM. The memorial symptom assessment scale: modified for use in understanding family caregivers' perceptions of cancer patients' symptom experiences. *J Pain Symptom Manage* 2003; 26: 644–654.
76. Chang VT, Hwang SS, Kasimis B, Thaler HT. Shorter symptom assessment instruments: the Condensed Memorial Symptom Assessment Scale (CMSAS). *Cancer Invest* 2004; 22: 526–536.
77. Hickman SE, Tilden VP, Tolle SW. Family reports of dying patients' distress: the adaptation of a research tool to assess global symptom distress in the last week of life. *J Pain Symptom Manage* 2001; 22: 565–574.
78. Byock IR, Merriman MP. Measuring quality of life for patients with terminal illness: the Missoula-VITAS quality of life index. *Palliat Med* 1998; 12: 231–244.
79. Schwartz CE, Merriman MP, Reed G, Byock I. Evaluation of the Missoula-VITAS Quality of Life Index – revised: research tool or clinical tool? *J Palliat Med* 2005; 8: 121–135.
80. Rainbird KJ, Perkins JJ, Sanson-Fisher RW. The Needs Assessment for Advanced Cancer Patients (NA-ACP): a measure of the perceived needs of patients with advanced, incurable cancer. a study of validity, reliability and acceptability. *Psycho-Oncology* 2005; 14: 297–306.
81. Vernooij-Dassen MJ, Osse BH, Schade E, Grol RP. Patient autonomy problems in palliative care: systematic development and evaluation of a questionnaire. *J Pain Symptom Manage* 2005; 30: 264–270.
82. Chochinov HM, Hassard T, McClement S, et al. The Patient Dignity Inventory: a novel way of measuring dignity related distress in palliative care. *J Pain Symptom Manage* 2008; 36: 559–571.
83. Osse BH, Vernooij MJ, Schade E, Grol RP. Towards a new clinical tool for needs assessment in the palliative care of cancer patients: the PNPC instrument. *J Pain Symptom Manage* 2004; 28: 329–341.
84. Osse BH, Vernooij-Dassen MJ, Schade E, Grol RP. A practical instrument to explore patients' needs in palliative care: the Problems and Needs in Palliative Care questionnaire short version. *Palliat Med* 2007; 21: 391–399.
85. Hearn J, Higginson IJ. Development and validation of a core outcome measure for palliative care: the Palliative Care Outcome Scale. *Qual Health Care* 1999; 8: 219–227.
86. Curtis JR, Patrick DL, Engelberg RA, Norris K, Asp C, Byock I. A measure of the quality of dying and death. Initial validation using after-death interviews with family members. *J Pain Symptom Manage* 2002; 24: 17–31.
87. Steinhauser KE, Clipp EC, Bosworth HB, et al. Measuring quality of life at the end of life: validation of the QUAL-E. *Palliat Support Care* 2004; 2: 3–14.
88. Hermann C. Development and testing of the spiritual needs inventory for patients near the end of life. *Oncol Nurs Forum* 2006; 33: 737–744.

Appendix 1. Search strategy

#1 (Palliative Care OR palliative OR Terminal Care OR terminal OR end of life OR limited life OR Hospice Care OR After-Hours Care)

 # 2 (Quality of Life OR quality of life) OR (Religion and Psychology OR spiritual*)

#1 AND #2→#3

#4 (addresses OR biography OR case reports OR comment OR directory OR editorial OR festschrift OR interview OR lectures OR legal cases OR legislation OR letter OR news OR newspaper article OR patient education handout OR popular works OR congresses OR consensus development conference OR consensus development conference, nih OR practice guideline) NOT (animals NOT humans)

#5 (Clinical Audit OR audit OR outcome assessment (health care) OR instrumentation OR Validation Studies OR reproducibility of results OR reproducib* OR psychometrics OR psychometr* OR clinimetr* OR clinometr* OR item selection OR item reduction OR observer variation OR observer variation OR discriminant analysis OR reliab* OR valid* OR coefficient OR internal consistency OR (cronbach* AND (alpha OR alphas)) OR item correlation OR item correlations OR item selection OR item selections OR item reduction OR item reductions OR agreement OR precision OR imprecision OR precise values OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobeserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappa's OR kappas OR coefficient of variation OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR known group OR factor analysis OR factor analyses OR factor structure OR factor structures OR dimensionality OR subscale* OR multitrait scaling analysis OR multitrait scaling analyses OR item discriminant OR interscale correlation OR interscale correlations OR ((error OR errors) AND (measure* OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR mean)) OR individual variability OR interval variability OR rate variability OR variability analysis OR (uncertainty AND (measurement OR measuring)) OR standard error of measurement OR sensitiv* OR responsive* OR (limit AND detection) OR minimal detectable concentration OR interpretab* OR (small* AND (real OR detectable) AND (change OR difference)) OR meaningful change OR minimal important change OR minimal important difference OR minimally important change OR minimally important difference OR minimal detectable change OR minimal detectable difference OR minimally detectable change OR minimally detectable difference OR minimal real change OR minimal real difference OR minimally real change OR minimally real difference OR ceiling effect OR floor effect OR Item response model OR IRT OR Rasch OR Differential item functioning OR DIF OR computer adaptive testing OR item bank OR cross-cultural equivalence)

(#3 NOT #4) AND #5

Appendix 2. Quality criteria for measurement properties

Property	Definition	Quality criteria[a,b]
Content validity	The extent to which the domain of interest is represented by the items in the questionnaire.	+ A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection AND a full copy of the instrument should be available; ? A clear description of above-mentioned aspects is lacking OR only target population involved OR doubtful design or method OR a full copy of the instrument is lacking; - No target population involvement; 0 No information found on target population involvement.
Construct validity	The extent to which scores on a particular instrument correspond to other measures in a manner that is consistent with theoretical expectations	+ Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses;

(continued)

Appendix 2. Continued

Property	Definition	Quality criteria[a,b]
		? Doubtful design or method (e.g. no hypotheses); - Less than 75% of hypotheses were confirmed, despite adequate design and methods; 0 No information found on construct validity.
Internal consistency	The extent to which items in a (sub) scale are intercorrelated, thus measuring the same construct.	+ Factor analyses performed on adequate sample size (7 * # items AND ≥100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95[c]; ? No factor analysis OR doubtful design; - Cronbach's alpha(s) <0.70 or >0.95, despite adequate design and method[d]; 0 No information found on internal consistency.
Reliability	The extent to which the instrument is able to distinguish patients from each other, despite measurement error (relative measurement error).	+ ICC or weighted Kappa ≥0.70 AND time interval at least 1 week[e]; ? Doubtful design or method (e.g. time interval not mentioned); - ICC or weighted Kappa <0.70, despite adequate design and method; 0 No information found on reliability.
Absolute measurement error	The absolute difference between two repeated measures.	+ SEM OR MIC <SDC or MIC outside the LOA OR convincing arguments that the measurement error is acceptable; ? Doubtful design of method (OR SEM or MIC not defined AND no convincing arguments that the measurement error is acceptable); - SDC or SDC ≥MIC or MIC equals or inside LOA OR RR ≤ 1.96 OR AUC < 0.70, despite adequate design and methods; 0 No information on absolute measurement error.
Responsiveness	The capacity of an instrument to detect clinically important changes over time.	+ Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses AND at least 2 measurements are available AND the time interval is described OR SDC or SDC<MIC or MIC outside the LOA OR RR>1.96 OR AUC ≥0.70; ? Doubtful design or method (e.g., no hypotheses); - Less than 75% of hypotheses were confirmed, despite adequate design and methods OR SDC or SDC≥MIC or MIC equals or inside LOA OR RR≤1.96 OR AUC <0.70, despite adequate design and methods; 0 No information on responsiveness.
Interpretability	The degree to which (change) scores can be interpreted.	+ Mean and SD scores presented of at least four relevant subgroups of patients and MIC defined and no floor/ceiling effects were present; ? Doubtful design of method OR less than four subgroups OR no MIC defined OR floor/ceiling effects were present; 0 No information found on interpretability.

ICC: intraclass correlation, SEM: standard error of measurement, MIC: minimal important change, SDC: smallest detectable change, LOA: limits of agreement, AUC: area under the curve, RR: responsiveness ratio.

[a]+: positive rating, ?: indeterminate rating, -: negative rating, 0: no information available.

[b]Doubtful design or method: lacking of a clear description of the design or methods or the study, sample size smaller than 50 subjects (should be at least 50 in every (subgroup) analysis), or any important methodological weakness in the design or execution of the study.

[c]75% of Cronbach's alphas between 0.70 and 0.90 AND no Cronbach's alpha <0.50.

[d]<75% of Cronbach's alphas between 0.70 and 0.90 OR Cronbach's alpha <0.50.

[e]Time interval at least 1 week OR less than 1 week when the questionnaire contains 30 items OR less than 1 week when convincing arguments were given that the time interval was appropriate.

Appendix 3. Full names of the questionnaires included

BHI	Brief Hospice Inventory[55]
CAMPAS-R	Cambridge Palliative Audit Schedule[56]
DS	Demoralization Scale[57]
EFAT	Edmonton Functional Assessment Tool[58,59]
	Emanuel and Emanuel Medical Directive[60]
EORTC QLQ-OES18	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Oesophageal cancer module[61]
EORTC QLQ-STO22	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Gastric cancer module[62]
ESAS	Edmonton Symptom Assessment Scale[63]
FACIT-Pal	Functional Assessment of Chronic Illness Therapy-Palliative sub scale[64]
HQLI	Hospice Quality of Life Index[65,66]
LCS	Life Closure Scale[67]
LEQ	Life Evaluation Questionnaire[68]
MQLS	McMaster Quality of Life Scale[69]
MQOL	McGill Quality of Life Questionnaire[70,71]
MQOL-CSF	McGill Quality of Life Questionnaire-Cardiff Short Form[72]
MRDI	McCanse Readiness for Death Instrument[73]
MSAS	Memorial Symptom Assessment Scale[74,75]
CMSAS	Condensed Memorial Symptom Assessment Scale[76]
MSAS-GDI	Memorial Symptom Assessment Scale-Global Distress Index[77]
MVQOLI	Missoula-VITAS Quality of Life Index[78,79]
NA-ACP	Needs Assessment for Advanced Cancer Patients[80]
PAQ	Patient Autonomy Questionnaire[81]
PDI	Patient Dignity Inventory[82]
PNPC	Problems and Needs in Palliative Care questionnaire[83]
PNPC-sv	Problems and Needs in Palliative Care questionnaire-short version[84]
POS	Palliative care Outcome Scale[85]
QODD	Quality of Dying and Death questionnaire[86]
QUAL-E	Quality of life at the end of life[87]
SNI	Spiritual Needs Inventory[88]