




Tools Measuring Quality of Death, Dying, and Care, Completed after Death: Systematic Review of Psychometric Properties

Nuriye Kupeli¹  · Bridget Candy¹ · Gabrielle Tamura-Rose² · Guy Schofield³ · Natalie Webber⁴ · Stephanie E. Hicks⁵ · Theodore Floyd⁶ · Bella Vivat¹ · Elizabeth L. Sampson^{1,7} · Patrick Stone¹ · Trefor Aspden⁸

Published online: 24 August 2018
© The Author(s) 2018

Abstract

Introduction Measuring the quality of care at the end of life and/or the quality of dying and death can be challenging. Some measurement tools seek to assess the quality of care immediately prior to death; others retrospectively assess, following death, the quality of end-of-life care. The comparative evaluation of the properties and application of the various instruments has been limited.

Objective This systematic review identified and critically appraised the psychometric properties and applicability of tools used after death.

Method We conducted a systematic review according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines by systematically searching MEDLINE, Embase, CINAHL, and PsycINFO for relevant studies. We then appraised the psychometric properties and the quality of reporting of the psychometric properties of the identified tools using the COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) checklist. The protocol of this systematic review has been registered on PROSPERO (CRD42016047296).

Results The search identified 4751 studies. Of these, 33 met the inclusion criteria, reporting on the psychometric properties of 67 tools. These tools measured quality of care at the end of life ($n = 35$), quality of dying and death ($n = 22$), or both quality of care at the end of life and dying and death ($n = 10$). Most tools were completed by family carers ($n = 57$), with some also completed by healthcare professionals (HCPs) ($n = 2$) or just HCPs ($n = 8$). No single tool was found to be adequate across all the psychometric properties assessed. Two quality of care at the end of life tools—Care of the Dying Evaluation and Satisfaction with Care at the End of Life in Dementia—had strong psychometric properties in most respects. Two tools assessing quality of dying and death—the Quality of Dying and Death and the newly developed Staff Perception of End of Life Experience—had limited to moderate evidence of good psychometric properties. Two tools assessing both quality of care and quality of dying and death—the Quality Of Dying in Long-Term Care for cognitively intact populations and Good Death Inventory (Korean version)—had the best psychometric properties.

Conclusion Four tools demonstrated some promise, but no single tool was consistent across all psychometric properties assessed. All tools identified would benefit from further psychometric testing.

The protocol of this systematic review has been registered on PROSPERO, which can be accessed here: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016047296 (Registration number: CRD42016047296).

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s40271-018-0328-2>) contains supplementary material, which is available to authorized users.

✉ Nuriye Kupeli
N.Kupeli@ucl.ac.uk

Extended author information available on the last page of the article

Key Points for Decision Makers

Psychometric information for measures assessing quality of dying and death and quality of care at the end of life was limited, so further research is required before a definitive choice of measure can be made.

Based on the limited evidence available, among the measures of quality of care at the end of life, the Care of the Dying Evaluation and Satisfaction with Care at the End of Life in Dementia tools appeared to have the best psychometric properties overall.

Among quality of dying and death measures, the Quality of Dying and Death and Staff Perception of End of Life Experience instruments appeared to have the best psychometric properties overall.

1 Introduction

By 2040, of people dying in England and Wales, 87.6% will need palliative care [1]. People at the end of life may experience difficult symptoms, such as pain, difficulty breathing, and confusion [2]. Multiple tools seek to assess the quality of care at the end of life and the quality of dying and death [3–7]. However, assessing the quality of dying and death and of end-of-life care can be challenging because of declining health towards the end of life, the difficulty of identifying people who may be in the dying phase, and the sensitivity of involving family members in quality assessment at this time. Additionally, development and validation of new tools is costly and time consuming. Thus, research might more productively evaluate and improve existing tools rather than developing new ones.

Tools that assess quality of care at the end of life reflect the provision of care and include items that assess the environment, communication with health and social care practitioners, and nursing care. In contrast, tools designed to assess quality of dying and death include items that reflect physical, psychological, emotional, and spiritual needs; symptom burden; and place of death. Several published systematic reviews have assessed the utility of these tools in various clinical populations, including dementia and cancer [3–7]. Notably, a recent review distinguished between tools assessing quality of dying and death and those examining quality of care in long-term care settings [7]. Similarly, van Soest-Poortvliet et al. [6] used a structured approach to assess the psychometric properties of tools developed to capture the quality of end-of-life care and of dying in long-term care settings and how they may differ for people with and without dementia. However, this group based their

assessment of the psychometric properties of tools in this field on data they collected in the USA and the Netherlands. The present review is broader. It assesses the psychometric properties of all tools developed and validated to evaluate, following death, the quality of dying and death and of care at the end of life across multiple settings and the methodological quality of the studies reporting these psychometric properties. Although retrospective recall of these concepts is susceptible to issues relating to recall bias, this approach overcomes the issue of whether or not the patient was at the end of life.

This review uses the COSMIN (consensus-based standards for the selection of health measurement instruments) [8], a taxonomy developed to standardise terminology and definitions of psychometric properties [9] and provide guidance on the best methods for developing and validating tools [10]. Since its development, the COSMIN [8] has been used to assess tools developed for various clinical populations, including dementia [11, 12] and breast cancer [13], and tools assessing quality of life in palliative care samples [14] and quality of care and dying in long-term care settings [6].

1.1 Aims

This systematic review aimed to (1) identify all tools that, after death, assess the quality of death and dying and of care at the end of life, (2) evaluate the psychometric properties of these tools, and (3) recommend validated tools for use in research and clinical practice.

2 Method

The protocol for this review is registered on PROSPERO (CRD42016047296) and follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15]. See the Electronic Supplementary Material (ESM) 1 for the PRISMA checklist.

2.1 Search Strategy

We searched the Cumulative Index to Nursing and Allied Health Literature (CINAHL), Embase, MEDLINE, and PsycINFO databases from inception to 15 May 2017 using search terms including medical subject heading (MeSH) terms for tools, end of life, quality of death and dying, and quality of care (Box 1). Terms for the end-of-life care population were extracted from a previous Cochrane review [16]. The reference lists of included studies and relevant reviews were examined to identify additional suitable studies.

Box 1: Search terms used

#1 (Tool or Assess* or Survey or Question* or Measur* or Method or Scale or Checklist or Rating or Test or Instru* or Inventory or Technique or Monitor or Observ* or Rate or Function or scoring system or Outcome*)

#2 (palliat* or terminal* or endstage or hospice* or (end adj3 life) or (care adj3 dying) or ((advanced or late or last or end or final) adj3 (stage* or phase*)))

#3 (Quality of death or Quality of dying or Satis* with care or Good death or Bad death or Comfort)

#4 1 and 2 and 3

2.2 Eligibility Criteria

Inclusion criteria were as follows: (1) studies assessed at least one psychometric property of a tool assessing quality of death and dying and/or of care at the end of life of adult palliative care patients in either inpatient or community settings, (2) tools were completed after death by family and healthcare professionals (HCPs) of deceased patients, (3) studies were reported in English even if psychometric properties of the tool were developed and validated in another language, and (4) studies were published in peer-reviewed journals. Exclusion criteria were studies that reported (1) ad hoc tools, (2) single-item tools, (3) tools developed for study purposes, and (4) tools developed for critical care settings (e.g., intensive care units). As the COSMIN was developed to assess the methodological quality of studies using classical test theory or item response theory (IRT), we excluded studies that used other methods such as generalisability theory [17]. Studies of assessment tools such as the Mini-Suffering State Examination (MSSE) [18] and Palliative Outcome Scale (POS) [19] were excluded because the MSSE assesses a variety of symptoms that are not designed to correlate and thus are not reflective of an overall construct. The POS has been shown to capture two factors and some independent items that do not load onto these factors [20], making this measure less ideal for the assessment of internal consistency and factor structure.

2.3 Study Selection

The study selection process consisted of two phases. First, review of all citations followed by a full-text review of studies that fitted the inclusion criteria according to this initial screening process. One researcher (NK) screened all titles and abstracts, and three reviewers (BV, JH, TA) each independently assessed a random sample of 250 abstracts and titles (750 in total). Any discrepancies were resolved through discussion and consensus. This process aimed to

facilitate clarity and agreement in the study group that our inclusion criteria was appropriate and sufficiently detailed to apply. One researcher (NK) screened all full-text studies and consulted with a second reviewer (TA) if the relevance of a study was unclear. Study authors were contacted if the relevance of a paper was unclear or additional information was required.

2.4 Data Extraction

Two reviewers (NK in all cases, plus one of TA, GTR, GS, NW, SH, and TF) independently extracted data from each study using a standardised data extraction form. Data extracted were country of origin, aim of study, tool(s) developed and/or validated, tool aim(s), number of items, response scale, language of tool assessed, respondent (informal [family] or paid [HCP] carer), recall period, method of administration, study setting, patient population, sample size, and sample demographic information of respondents and/or of the deceased patients.

2.5 Assessment of Psychometric Properties

Psychometric properties of tools were appraised using established quality criteria ([21, 22]; Table 1). The COSMIN provides guidance for assessing a range of psychometric properties, including validity (content, construct [structural, hypothesis testing, and cross-cultural], and criterion), reliability (internal consistency, reproducibility [agreement and reliability over time and between and within raters], responsiveness and floor and ceiling effects). To our knowledge, there is no ‘gold standard’ tool for measuring the quality of care at the end of life and quality of dying, and thus criterion validity was not assessed. Each psychometric property was scored using a four-point rating scale: positive (+), indeterminate (?), negative (−), or no information (0). The criteria for the positive, indeterminate, and negative ratings for each of the psychometric properties assessed in this review are presented in Table 1.

Table 1 Quality criteria used to assess psychometric properties of measures [22]

Psychometric property	Definition	Rating	Quality criteria
Internal consistency	The extent to which the items correlate, indicating that the overall tool is measuring the same construct	+	Data from adequate sample used to conduct factor analysis and Cronbach's α per factor between 0.70 and 0.95
		?	No factor analysis conducted
		-	Cronbach's $\alpha < 0.70$ or > 0.95
Reliability	The degree to which the scores are free from measurement error	+	ICC or weighted kappa ≥ 0.70
		?	ICC or weighted kappa not reported or inappropriate statistical method reported
		-	ICC or weighted kappa < 0.70
Content validity	The extent to which the items reflect the construct being assessed	+	Detailed description of tool development, including tool aim, target population, concepts under assessment, item selection, and the population (patient and experts) involved in item selection
		?	Aspects of tool development lack description or only target population involved
		-	No target population involved
Structural validity	The degree to which the scores of the tool adequately reflect the dimensions of the construct being assessed	+	Factor analysis demonstrates that combined set of factors explain $\geq 50\%$ of total variance in the model OR IRT methods confirm (uni) dimensionality
		?	Proportion of variance explained not reported
		-	$< 50\%$ of the total variance explained by model
Hypothesis testing	The extent to which the scores of the tool are consistent with pre-formulated hypotheses	+	Specific hypotheses were formulated and at least 75% of the results in line with hypotheses
		?	No a priori formulated hypotheses
		-	Results confirm $< 75\%$ of hypotheses
Cross-cultural validity	The degree to which the items on a translated version are an adequate reflection of the original version		Assessed using methodological quality criteria only
Item response theory	Assesses whether responses on a set of items are related to an unmeasured 'trait'		Assessed using methodological quality criteria only

Psychometric property ratings: + indicates positive; ? indicates indeterminate; - indicates negative

ICC Intraclass correlation coefficient, IRT item response theory

2.6 Assessment of Methodological Quality

We used the COSMIN checklist to appraise the methodological quality of studies reporting on psychometric properties of the tools [21]. This checklist comprises nine boxes, each rating a specific psychometric property. Each psychometric property is rated on 5–18 items as excellent, good, fair, or poor. Methodological and psychometric quality were assessed for all psychometric properties except cross-cultural validity and IRT, which were only rated on methodological quality. Appraisal of each measure was based on the overall tool where possible; however, for studies that reported the psychometric properties of individual subscales rather than the overall tool, these scales were assessed individually. Assessment of psychometric property and methodological quality for each study was completed by two independent reviewers (NK for all studies, plus one of TA, GTR, GS, NW, SH, and TF). Each rating was compared and any discrepancies between the two reviewers discussed

and resolved, with a third rater consulted if no resolution could be reached. Intraclass correlation coefficients (ICCs) between reviewers for the assessment of methodological quality of each psychometric property ranged between 0.70 and 0.97, and high agreement was found for psychometric property appraisal (ICC range 0.87–1.0).

2.7 Levels of Evidence

The psychometric property assessed for each measure was accompanied by an assessment of the level of evidence available to support the rating. The level of evidence was determined by the number of studies reporting on the psychometric property of the measure, the methodological quality as assessed by the COSMIN, and the agreement between studies if more than one had been conducted. Each psychometric property was rated either as strong (consistent findings across several studies with a methodological rating of 'good' or one study rated as 'excellent'), moderate

(consistent evidence across several studies rated as ‘fair’ or one study rated as ‘good’ in methodological quality), limited (findings from one study rated as ‘fair’), unknown (findings from studies rated as ‘poor’ available), or conflicting (inconsistent findings across different studies) [23].

2.8 Data Synthesis

Rating data from different studies evaluating the same tools were grouped based on the methodology. Data from studies that used the same version of the tool (i.e. on the same items, response scale, and language) and collected data from the same types of respondents (i.e. family carers or HCPs) were suitable for grouping. For tools where it was not possible to group the data from two or more studies, the ratings for each study were presented individually. For tools where data grouping was possible, only data for the psychometric properties rated as fair, good, or excellent on methodological quality as assessed by the COSMIN [21] were used.

We used the COSMIN [21] to assess the psychometric properties of each tool and the methodological quality of reporting, but, to make a global comparison between the tools, we developed an additional ad hoc scoring system, assigning a score for the psychometric property rating and level of evidence for each psychometric property assessed (Box 2).

Box 2: Scoring system used to compare overall properties of the tools

Psychometric property rating	Level of evidence	Score assigned
+	Strong	+3
+	Moderate	+2
+	Limited	+1
-	Limited	-1
-	Moderate	-2
-	Strong	-3

Psychometric properties that were rated as indeterminate (?), unknown, or conflicting were assigned a score of 0. The scores assigned for each psychometric property were summed to give an overall score for each tool.

3 Results

3.1 Search Results and Study Selection

A total of 4751 studies were retrieved from the database searches. Following screening of abstracts and titles, 347 studies were taken forward for full-text review, and 28 studies fitted the inclusion criteria. Reference list checks of the 28 relevant studies identified an additional five relevant

studies, resulting in a final list of 33 studies to be included in the review. A PRISMA flow diagram of the screening process is presented in Fig. 1.

In total, the 33 studies assessed 67 tools. The majority of the studies assessed the psychometric properties of tools completed by family carers ($n=57$), but some were completed by HCPs ($n=8$) or both ($n=2$). The tools were completed in English ($n=44$), Dutch ($n=11$), German ($n=2$), Japanese ($n=4$), Korean ($n=2$), Spanish ($n=2$), and Italian ($n=1$). Another study used both English and Spanish versions of a tool. Studies were conducted in the USA ($n=14$), Japan ($n=4$), UK ($n=3$), Netherlands ($n=3$), Korea ($n=2$), and Germany ($n=2$). One study was international, with participants from Canada, Chile, Ireland, Italy, and Norway. Studies evaluated the quality of care and of dying and death in palliative care units ($n=10$); long-term care settings, including nursing homes ($n=7$), hospitals ($n=2$), hospices ($n=1$), home ($n=1$), outpatient units ($n=1$); and across various settings ($n=9$). The clinical populations were from a mixed sample ($n=15$) or had advanced cancer ($n=14$) or advanced dementia ($n=4$).

All but one study [7] evaluated the psychometric properties of the overall tool and/or the subscales of the tool except for the minimum data set (MDS), which was evaluated as individual subscales rather than an overall tool. ESM 2 provides a summary of the included studies. Tools completed by family carers and HCPs, translated versions, and tools evaluated as independent subscales were evaluated individually, providing assessments of 67 tools. Of these tools, 35 assessed quality of care, 22 measured quality of dying and death, and ten assessed both. While the majority of studies assessed the psychometric properties of one version of a measure completed by a single sample ($n=21$), some assessed two ($n=8$), three ($n=1$), four ($n=1$), 11 ($n=1$), or 12 ($n=1$) individual tools that differed either in what they evaluated or in the respondent who completed the measure (family carer or HCPs). ESM 3 provides a summary of all tools used in each included study.

3.2 Psychometric Properties of Tools

Data on psychometric properties for all of the tools could not be grouped because of substantial differences between studies in the versions of the tools used (i.e. original, abbreviated, different language), the method of using the tools (i.e. family carers and/or HCPs, self-administered and/or interview), and the settings (i.e. long-term care, hospice, hospital). Therefore, rating data on tools from two or more studies that could not be grouped are presented individually, whereas studies that used the same tools with similar methodology have been grouped. Table 2 provides a summary of the rating assigned for each psychometric property, the level of evidence, and the overall score for each tool, and ESM

Fig. 1 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study selection. *EOLC* end-of-life care, *ICU* intensive care unit

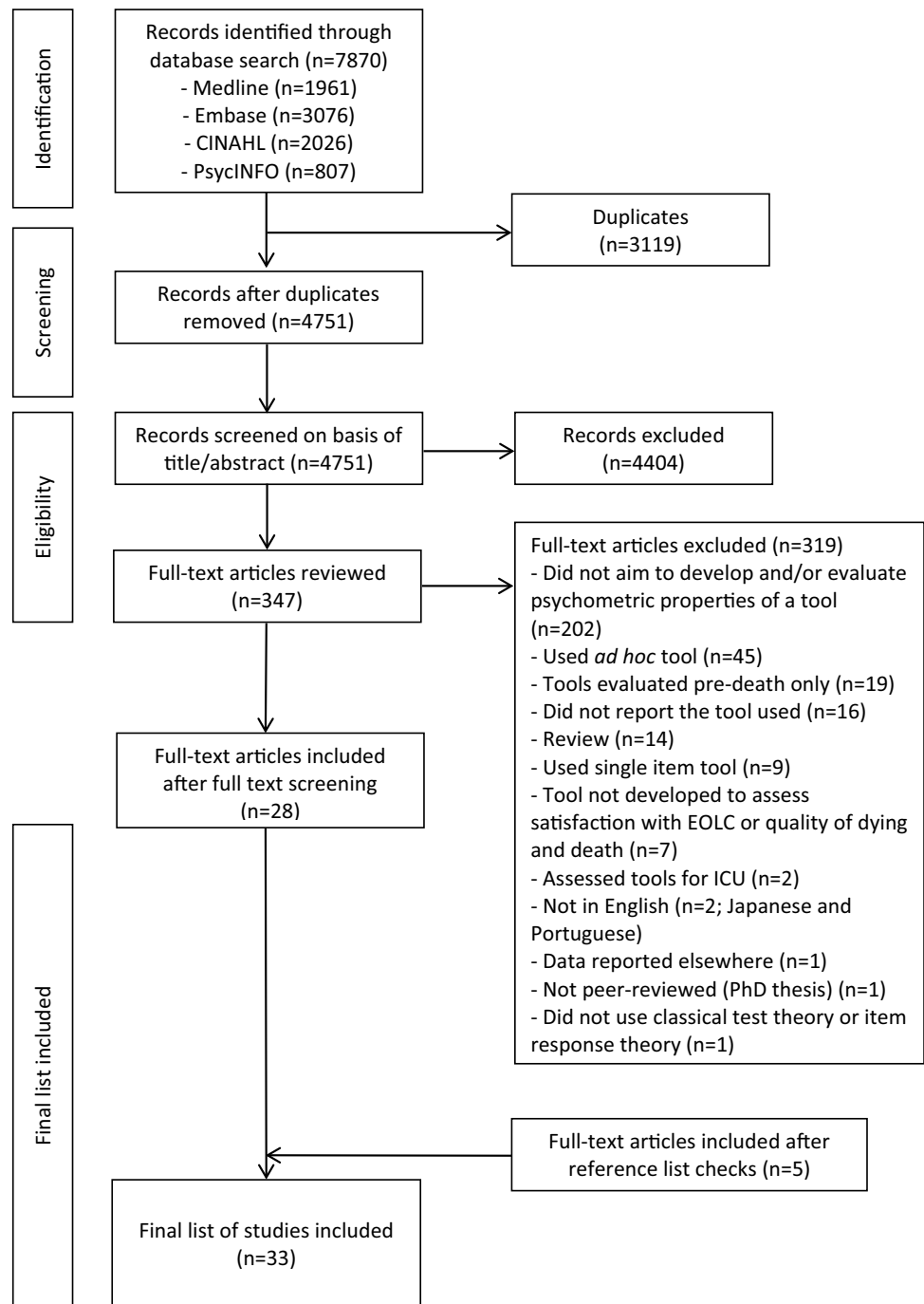


Table 2 Data synthesis of quality of psychometric properties and level of evidence for tools

Tool (respondent)	Studies (n)	Internal consistency		Reliability		Content validity		Structural validity		Hypotheses testing		Cross-cultural		IRT	Score
		PQ	LOE	PQ	LOE	PQ	LOE	PQ	LOE	PQ	LOE	PQ	LOE		
<i>Quality of care</i>															
CODE (FC) [31]	1	+	Moderate	Unknown	Unknown	+	Strong	?	Moderate	?	Limited				5
SWC-EOLD (FC) [7, 24]	2	+	Strong	Unknown	Unknown	+	Unknown	?	Strong	?	Moderate				3
FATE-32 (FC) [29]	1	+	Limited	Limited	Limited	+	Limited	?	Limited	?	Limited				2
CQ-Index-PC ² (FC) [33]	1	?	Limited	Limited	Limited	+	Moderate	?	Limited	?	Limited				2
QPM-SF ³ (FC) [34]	1	-	Limited	-	Limited	+	Strong	+	Limited	Unknown	Unknown				2
FAMCARE-10 (FC) [30]	1	+	Limited	Limited	Limited	+	Limited	+	Limited					Moderate	2
FAMCARE-5 (FC) [30]	1	+	Limited	Limited	Limited	+	Limited	+	Limited					Moderate	2
FPCS (FC) [7, 25]	2	-	Limited	Limited	Moderate	+	Moderate	+	Limited	?	Moderate				2
TIME (FC) [7, 26]	2	+	Moderate	Unknown	Unknown	+	Moderate	?	Moderate	?	Moderate				2
SAT-Fam-IPC (FC) [35]	1	-	Moderate	Moderate	Moderate	+	Moderate	+	Moderate	Unknown	Unknown				2
CES-10 ⁴ (FC) [32]	1	Unknown	Unknown	+	Moderate										2
FPPFC (FC) [7]	1	+	Moderate	Moderate	Moderate			?	Moderate	?	Limited				2
FAMCARE (FC) [37]	1	-	Moderate	Moderate	Moderate	+	Strong	+	Strong					Moderate	1
ECHO-D (FC) [27, 28]	2	+	Limited	-	Limited			Unknown	Unknown	+	Limited				1
MDS-Mood (FC) [7]	1	+	Limited	Limited	Limited			?	Moderate	?	Limited				1
FATE-S-14 ¹ (FC) [38]	1	Unknown	Unknown	Unknown	Unknown					?	Limited			Unknown	0
FATE-S-12 (FC) [7]	1	Unknown	Unknown	Unknown	Unknown					?	Limited			Unknown	0
FATE-S-12 ² (FC) [39]	1	Unknown	Unknown	Unknown	Unknown					Unknown	Limited			Unknown	0
SWC-EOLD ² (FC) [39]	1	Unknown	Unknown	Unknown	Unknown					?	Limited			Unknown	0
FPCS ² (FC) [39]	1	Unknown	Unknown	Unknown	Unknown					Unknown	Limited			Unknown	0
TIME ² (FC) [39]	1	Unknown	Unknown	Unknown	Unknown					?	Limited			Unknown	0
McCusker EOLC scale (FC) [54]	1	Unknown	Unknown	Unknown	Unknown					Unknown	Limited			Unknown	0
CES (FC) [36]	1	-	Limited	-	Limited	+	Moderate	?	Limited	Unknown	Unknown			Unknown	0
CES ⁴ (FC) [32]	1	-	Moderate	+	Moderate			?	Moderate	?	Limited			Unknown	0
FPPFC ² (FC) [39]	1	Unknown	Unknown	Unknown	Unknown					?	Limited			Unknown	0
MDS-Spirituality (FC) [7]	1	Unknown	Unknown	Unknown	Unknown					?	Limited			Unknown	0
MDS-Social (FC) [7]	1	-	Limited	Limited	Limited			?	Moderate	?	Limited				-1
MDS-Symptoms (FC) [7]	1	-	Moderate	Moderate	Moderate			?	Moderate	?	Limited				-2
CES ⁵ (FC) [40]	1	-	Strong	Strong	Strong			?	Strong	Unknown	Limited			Limited	-3
CEQUEL (FC) [41]	1	-	Strong	Strong	Strong			?	Strong	?	Limited			Limited	-3
<i>Quality of dying and death</i>															
SPELE (HCPs) [47]	1	Unknown	Unknown	Unknown	Unknown	+	Moderate	?	Moderate						2
QODD (FC) [42, 43]	2 ^a	Unknown	Unknown	Unknown	Unknown			?	Strong	+	Limited				1
CAD-EOLD (FC) [7, 24]	2	Conflicting	Conflicting	Conflicting	Conflicting	Unknown	Unknown	Conflicting	Conflicting	?	Moderate				0

Table 2 (continued)

Tool (respondent)	Studies (n)	Internal consistency		Reliability		Content validity		Structural validity		Hypotheses testing		Cross-cultural		IRT	Score
		PQ	LOE	PQ	LOE	PQ	LOE	PQ	LOE	PQ	LOE	PQ	LOE		
CAD-EOLD ² (FC) [39]	1	?	Limited	Unknown	Unknown					?	Limited				0
CAD-EOLD (HCPs) [55]	1	Unknown		Unknown						?	Limited				0
SM-EOLD (FC) [7, 24]	2	Conflicting		Unknown	Unknown			Conflicting		?	Moderate				0
SM-EOLD ² (FC) [39]	1			Unknown						?	Limited				0
SM-EOLD ² (HCPs) [39]	1			Unknown						?	Limited				0
SM-EOLD (HCPs) [56]	1	Unknown		Unknown						?	Limited				0
MSAS-GDI (FC) [57]	1	Unknown		Unknown						?	Limited				0
QODD-ESP ⁷ (FC) [46]	1	Unknown		Unknown						Unknown			Limited		0
QODD-ESP-12 ⁷ (FC) [46]	1			Unknown				?	Moderate	Unknown			Limited		0
CAD-EOLD ² (HCPs) [39]	1	-	Limited	Unknown				?	Limited	?	Limited				-1
QODD-D-Ang ⁶ (FC) [44]	1	Unknown		?	Limited	Unknown		-	Strong	?	Limited		Unknown		-3
QODD-D-MA ⁶ (HCPs) [45]	1	Unknown		?	Limited			-	Strong	Unknown			Unknown		-3
<i>Quality of care and dying and death</i>															
QOD-LTC-C (FC and HCPs) [48]	1	+	Strong	Unknown				?	Strong						3
GDI ⁵ (FC) [49]	1	+	Strong	Unknown				?	Strong	Unknown			Limited		3
GDI-Short version ⁴ (FC) [51]	1	Unknown		Unknown						Unknown			Limited		0
QOD-Hospice Short Form (FC) [50]	1	Unknown		Unknown				Unknown		Unknown					0
QOD-LTC (FC) [7]	1	Unknown		Unknown				?	Moderate						0
QOD-LTC ² (FC) [39]	1	?	Limited	Unknown						?	Limited		Unknown		0
QOD-LTC ² (HCPs) [39]	1	-	Limited	Unknown				?	Limited	?	Limited		Unknown		-1
QOD-Hospice (FC) [50]	1	Unknown		-	Limited	Unknown		Unknown		-	Limited		Unknown		-2
GDI ⁴ (FC) [51]	1	Unknown		-	Moderate	?	Limited			?	Limited		Unknown		-2
QOD-LTC (FC and HCPs) [48]	1	-	Strong	-	Moderate	Unknown		-	Strong				Strong		-8

Psychometric property quality ratings: + indicates positive, ? indicates indeterminate, - indicates negative

Level of evidence: strong = consistent findings across several studies with a methodological rating of 'good' or one study rated as 'excellent'; moderate = consistent evidence across several studies rated as 'fair' or one study rated as 'good' in methodological quality; limited = findings from one study rated as 'fair'; unknown = findings from studies rated as 'poor' available; conflicting = inconsistent findings across different studies

Tool completed in: ¹ = English and Spanish; ² = Dutch; ³ = Italian; ⁴ = Japanese; ⁵ = Korean; ⁶ = German; ⁷ = Spanish

CAD-EOLD Comfort Assessment in Dying at the End of Life in Dementia, CEQUEL Caregiver Evaluation of the Quality of End of Life care, CES Care Evaluation Scale, CODE Caring Of the Dying Evaluation, CQ-Index-PC Consumer Quality Index Palliative Care, ECHO-D Evaluating Care and Health Outcomes for the Dying, EOLC end-of-life care, FAMCARE Family satisfaction

Table 2 (continued)

with end-of-life Care, *FATE* Family Assessment of Treatment at the End of life, *FC* family carers, *FPPFC* Family Perceptions of Physician-Family Caregiver Communication, *GDI* Good Death Inventory, *HCPs* healthcare professionals, *IRT* item response theory, *LOE* level of evidence, *MDS* minimum data set, *MSAS-GDI* Memorial Symptom Assessment Scale Global Distress Index, *PQ* psychometric quality, *QODD* Quality of Dying and Death, *QODD-D-Ang* *QODD*-Deutsch-Angehörige, *QODD-D-MA* *QODD*-Deutsch-Mitarbeiter, *QODD-Exp* Spanish version of the *QODD*, *QOD-Hospice* Quality of Dying-Hospice scale, *QOD-LTC* Quality of Dying in Long-Term Care, *QOD-LTC-C* Quality Of Dying in Long-Term Care of Cognitively intact decedents, *QPM-SF* Post Mortem Questionnaire-Short Form, *SAT-Fam-IPC* Satisfaction Scale for Family members receiving Inpatient Palliative Care, *SM-EOLD* Symptom Management at the End of Life in Dementia, *SPELE* Staff Perception of End of Life Experience, *SWC-EOLD* Satisfaction With Care at the End of Life in Dementia, *TIME* Toolkit of Instruments to Measure End of life care after-death bereaved family member interview

^aStudy 46 assessed the *QODD* in two separate family carer samples

4 provides the appraisal of the psychometric properties of tools and the methodological quality of studies using the COSMIN checklist.

3.2.1 Psychometric Properties of Tools Assessing Quality of Care at the End of Life

The tools identified to assess quality of care at the end of life are shown in Table 2. From these tools, it was possible to group the data from two studies assessing the *SWC-EOLD* (Satisfaction With Care at the End of Life in Dementia) [7, 24], *FPCS* (Family Perceptions of Care Scale) [7, 25], *TIME* (Toolkit of Instruments to Measure End of life care after-death bereaved family member interview) [7, 26], and *ECHO-D* (Evaluating Care and Health Outcomes-for the Dying) [27, 28]. Internal consistency, structural validity, and hypothesis testing were assessed for all four tools, whereas content validity was evaluated for *SWC-EOLD* [7, 24] and *FPCS* [7, 25], and reliability was assessed for *TIME* [7, 26] and the *ECHO-D* [27, 28]. The *SWC-EOLD* [7, 24] had strong evidence of positive internal consistency, with Cronbach's alpha (α) ranging between 0.83 and 0.90, but a moderate to strong level of evidence of an indeterminate rating for structural validity and hypothesis testing and an indeterminate rating from one study of poor methodological quality (thus rated as unknown) for content validity. In contrast, limited evidence showed that the *FPCS* [7, 25] had a negative rating for internal consistency ($\alpha = 0.95$ and 0.96) but a positive rating for content and structural validity and an indeterminate rating for hypotheses testing. The *TIME* [7, 26] measure had moderate evidence for positive internal consistency ($\alpha = 0.94$) but was indeterminate for structural validity and hypothesis testing and unknown for reliability. The *ECHO-D* [27, 28] also had limited evidence of positive internal consistency ($\alpha = 0.78$ – 0.93) for the subscales and was suitable for hypothesis testing but scored negatively for test–retest reliability ($\kappa < 0.70$).

For tools where data grouping was not possible, the *FATE* (Family Assessment of Treatment at the End of life)-32 [29], *FAMCARE* (Family satisfaction with end-of-life Care)-5 and -10 [30], *CODE* (Caring Of the Dying Evaluation) [31], *FPPFC* (Family Perceptions of Physician-Family Caregiver Communication) [7], and *MDS-Mood* [7] were all assigned a positive rating for internal consistency ($\alpha = 0.74$ – 0.94) but with varying levels of evidence. For reliability, the Japanese versions of the *CES* (Care Evaluation Scale) and *CES-10* [32] had moderate levels of evidence for positive test–retest reliability ($ICC = 0.82$ – 0.83). *FATE-32* [29], *CQ-Index-PC* (Consumer Quality Index Palliative Care) [33], *QPM-SF* (Post Mortem Questionnaire-Short Form) [34], *SAT-Fam-IPC* (Satisfaction Scale for Family members receiving Inpatient Palliative Care) [35], *CES* [36], and *CODE* [31] all had

strong to moderate levels of evidence of positive content validity, with strong evidence for QPM-SF [34] and CODE [31]. QPM-SF [34], FAMCARE [37], FAMCARE-5 and FAMCARE-10 [30], and SAT-Fam-IPC [35] all had positive structural validity properties, with a strong level of evidence for FAMCARE [37]. Cross-cultural validity was assessed for FATE-S-14 [38]; Dutch versions of the FATE-S-12, FPCS, TIME, and FPPFC [39]; SAT-Fam-IPC [35]; and the Korean and English versions of the CES [36, 40]. However, all studies were rated as poor for methodological quality except for the Korean version of the CES [40], which was rated as fair. Thus, the cross-cultural validity for the majority of tools is unknown (or limited in the case of the Korean version of the CES [40]). Although the CES [40] was rated as good to excellent on the majority of the methodological quality criteria, the authors failed to describe the expertise of the translators with respect to disease, construct, and language, and whether the translators worked independently was unclear. IRT methodology was used to assess the three versions of the FAMCARE scales (FAMCARE [37], FAMCARE-10 and FAMCARE-5 [30]); based on methodology, these were rated as good with moderate levels of evidence.

Using our ad hoc scoring system to assign an overall score for each tool, 15 of the 30 tools were assigned a positive score, with the CODE [31] and SWC-EOLD [7, 24] assigned a score $\geq +3$, whereas the MDS-Social [7], MDS-Symptoms [7], the Korean version of the CES [40], and the CEQUEL (Caregiver Evaluation of the Quality of End of Life care) [41] scored poorly (-1 to -3).

3.2.2 Psychometric Properties of Tools Assessing Quality of Dying and Death

The tools assessing quality of dying and death are shown in Table 2. Of these tools, it was possible to group data from studies with family carers assessing the CAD-EOLD (Comfort Assessment in Dying at the End of Life in Dementia), the SM-EOLD (Symptom Management at the End of Life in Dementia) [7, 24], and the QODD (Quality of Dying and Death) [42, 43]. Assessment outcomes of internal consistency for these tools were either conflicting (CAD-EOLD and SM-EOLD [7, 24]) or unknown (QODD [42, 43]). Although Cronbach's α for CAD-EOLD was within an acceptable range ($\alpha=0.74$ – 0.85), one study [24] did not employ an adequate sample (<100). The internal consistency evaluation of the SM-EOLD in two studies also differed, with one reporting a Cronbach's α of 0.72 [7]; although the overall scale internal consistency reported from the second study was within an acceptable range ($\alpha=0.78$), the subscales were found to have Cronbach's α of 0.47–0.81 [24]. One study assessed the internal consistency for the QODD, but this was rated as unknown because the factor structure was not evaluated [42]. Content validity was assessed for CAD-EOLD and

SM-EOLD for family carers [24] but was rated unknown because of the poor level of evidence. The QODD had a strong level of evidence for structural validity from two samples employed by one study [43], but this study failed to report the proportion of variance explained by the factorial models and thus was rated as indeterminate. The data for structural validity of the CAD-EOLD and SM-EOLD [7, 24] were conflicting. The assessment of hypothesis testing of the CAD-EOLD and SM-EOLD [7, 24] were both indeterminate because hypotheses were lacking. QODD was found to have positive hypothesis-testing properties, as one study formulated and presented specific hypotheses and at least 75% of the results were in line with the hypotheses [42].

For tools where data could not be grouped, the majority were rated as unknown or indeterminate on psychometric assessment. The Dutch version of the CAD-EOLD for HCPs [39] had negative internal consistency (α for subscales ranged between 0.64 and 0.89). Similarly, the German versions of the QODD for family carers and HCPs (QODD-D-Ang [QODD-Deutsch-Angehörige] [44] and the QODD-D-MA [QODD-Deutsch-Mitarbeiter] [45], respectively) both had negative structural validity because factor analysis demonstrated that all the factors together explained $<50\%$ of the total variance (QODD-D-Ang = 44.97%; QODD-D-MA = 43.8%). Cross-cultural validity was assessed for the QODD-D-Ang [44], QODD-ESP (Spanish version) [46], and QODD-D-MA [45]. Although both the QODD-D-Ang [44] and QODD-D-MA [45] were rated as excellent for the majority of criteria, neither study performed a confirmatory factor analysis. A confirmatory factor analysis is required to test for differences between the original and translated versions of the tool and to identify whether any items do not load on the original factor structure, suggesting that the items have a different meaning in the translated version. In contrast, a confirmatory factor analysis was conducted to test the factor structure of the QODD-ESP [46], but this tool was rated as limited because it scored as fair on several criteria. The authors failed to describe the expertise of the translators with respect to disease, construct, and language, it was unclear whether the translators worked independently, and only one forward and one backward translation of the items was conducted.

Using our ad hoc scoring system, of the 15 tools, only the SPELE (Staff Perception of End of Life Experience) for HCPs [47] and the QODD for family carers [42, 43] had a positive score but with a moderate to limited level of evidence. In contrast, the Dutch version of the CAD-EOLD for HCPs [39], the QODD-D-Ang for family carers [44], and the QODD-D-MA for HCPs [45] were rated negatively.

3.2.3 Psychometric Properties of Tools Assessing Both Quality of Care at the End of Life and Quality of Dying and Death

The tools identified to assess both quality of care at the end of life and quality of dying and death are shown in Table 2. We found substantial differences between studies assessing the same tools, so it was not possible to group the data for these tools. Internal consistency was positive for the QOD-LTC-C (Quality Of Dying in Long-Term Care of Cognitively intact decedents) [48] as completed by both family carers and HCPs and the Korean version of the GDI (Good Death Inventory) for family carers [49] ($\alpha=0.85$ and 0.93 , respectively). However, internal consistency was negative for QOD-LTC (Quality Of Dying in Long-Term Care) as completed by family carers and HCPs [48] and the Dutch version for HCPs [39]. Cronbach's α for the subscales ranged from 0.49 to 0.66 and from 0.37 to 0.75, respectively. Interrater reliability was negative for the QOD-Hospice (Quality Of Dying-Hospice scale) [50] and the QOD-LTC for family carers and HCPs [48], and the Japanese version of the GDI for family carers [51] had negative test-retest reliability. The authors reported ICC values of 0.49, 0.35, and 0.52, respectively. Where an assessment of structural validity was available, the tools were rated as unknown or indeterminate, except for the QOD-LTC for both family carers and HCPs [48], which was rated negative. The factor analysis found that the model explained 49% of the total variance. Although the authors formulated and reported specific hypotheses for the QOD-Hospice, the hypotheses testing for this tool was rated as negative because the results were not in line with at least 75% of the hypotheses [50]. Cross-cultural validity was assessed for Dutch versions of the QOD-LTC for family carers and HCPs [39] (rated as unknown) and the Korean version of the GDI [49] (rated as limited). The GDI [49] was rated as limited because, similar to other tools assessed in this study, the authors failed to report the expertise of the translators and did not clearly describe whether the translators worked independently.

Using our ad hoc scoring system, two of the ten tools (QOD-LTC-C for family carers and HCPs [48] and the Korean version of the GDI for family carers [49]) were rated positively and had a strong level of evidence. In contrast, four tools (the Dutch version of the QOD-LTC for HCPs [39], QOD-Hospice for family carers [50], the Japanese version of the GDI for family carers [51], and the QOD-LTC for family carers and HCPs [48]) were rated negatively, with the QOD-LTC for family carers and HCPs [48] assigned a score of -8 .

4 Discussion

4.1 Findings

To our knowledge, this is the first systematic review to identify and appraise psychometric properties while considering the associated levels of evidence for tools that, after death, assess the quality of care at the end of life and the quality of dying and death. Our review identified 33 studies that reported on versions of 35 tools assessing quality of care at the end of life, 22 tools assessing quality of dying and death, and ten assessing both constructs. Data on psychometric properties could not be grouped for every measure because of the variability between studies in the versions of tools used (i.e. original, abbreviated, different language), the method of using the tools (i.e. family carers and/or HCPs, self-administered and/or interview), and the settings (i.e. long-term care, hospice, hospital). However, no measure was adequate across all psychometric properties.

Our ad hoc scoring system rated half of the tools designed to assess quality of care positively. In particular, the CODE [31] which, although not psychometrically evaluated since it was developed, was initially assessed on five psychometric properties, with overall strong evidence of positive measurement properties. The CODE is a 30-item self-report measure developed from the ECHO-D [27, 28]. This tool is designed to assess the environment of the care setting, communication with HCPs, and the care provided to the patient in the last days of life. Despite its limited use, the CODE has some promising psychometric properties and thus should be developed and validated further. Another tool that also demonstrated strong evidence of positive psychometric properties, including internal consistency, is the SWC-EOLD [7, 24]. This tool is predominately used in long-term care settings to evaluate carers' satisfaction with end-of-life care provided to people living with dementia in the last 90 days of life. This is a 10-item self-report tool designed to assess decision making, communication with HCPs, understanding of dementia, and level of nursing care. Despite its extensive use in research, the SWC-EOLC would benefit from further psychometric evaluation, particularly, structural validity and hypotheses testing.

In contrast, the Korean version of the CES [40] and the CEQUEL [41] had strong evidence for negative and indeterminate ratings, suggesting that, to date, these tools have poor psychometric properties and thus require further development and validation. The majority of the psychometric properties of the tools developed to assess the quality of dying and death were rated as unknown or that the available evidence was conflicting, thus making it challenging to arrive at a firm conclusion on their

psychometric properties. For example, the majority of studies that assessed cross-cultural validity failed to adequately describe the translators with respect to their expertise on dying, death, and satisfaction with care, and the language, and whether the translators worked independently while translating the items was unclear.

On a positive note, the newly developed SPELE [47], which has been assessed for structural validity and content validity, had a moderate level of evidence of positive psychometric properties. The SPELE is a comprehensive 63-item tool designed to assess HCP's experiences of various aspects of quality of dying and death, including the environment, symptoms, decision making, and communication in the last week of life. This promising tool, although yet to be developed further, can be used across a variety of healthcare settings. In comparison, the QODD [42, 43], which was adapted from the original version and has been extensively used, also has some positive psychometric qualities. This 31-item tool measures a number of factors, including preparation for death, moment of death, and treatment preferences. As demonstrated by this review, the QODD has been translated into German and Spanish and used by both family carers and HCPs. However, despite extensive use, it still requires further validation, particularly for internal consistency and reliability.

Finally, of the ten tools designed to assess the quality of care and of dying and death, only the QOD-LTC for cognitively intact samples [48] and the Korean version of the GDI [49] had positive psychometric properties. Overall, the findings demonstrate that, of the numerous tools available to assess the quality of care and of dying and death, none have undergone a full psychometric evaluation with all psychometric properties evaluated. Further psychometric evaluation of the tools identified and assessed in the present review is required.

4.2 Strengths and Limitations

This systematic review can be considered of high methodological quality according to the quality criteria for systematic reviews proposed by Terwee et al. [52]. It used a comprehensive but broad search strategy without date restrictions to seek to capture all relevant articles from several key large citation databases and reference lists of suitable studies. Search terms for measurement properties were not used because of the great variation in the terminology, as recommended by the developers of the COSMIN [52]. A proportion of relevant studies were only identified through reference list checks, because the broad search strategy used when searching electronic databases still required studies to be correctly indexed and for appropriate keywords to be included in the title/abstract, but this was not always the case. A single reviewer (NK) assessed all of the results of

the search, but three secondary reviewers independently assessed a random sample of 750 titles and abstracts, and any discrepancies were resolved by discussion.

Only studies that aimed to develop and/or validate tools assessing, following death, the quality of care at the end of life and of dying and death were included in this review, so studies that assessed psychometric properties of tools as a secondary aim (i.e. it is common for research studies to include Cronbach's α of tools as an assessment of internal consistency) were excluded. Additionally, we included studies that reported psychometric properties of tools developed or validated in other languages to identify cross-cultural psychometric evidence for the tools of interest. Thus, this review is not restricted to English language tools or English-speaking populations and cultures.

The use of a well-defined and structured quality assessment such as the COSMIN provides a rigorous approach to psychometric evaluation. However, the COSMIN is not suitable for assessing studies that have used other methods, such as generalisability theory, as it was developed to assess the methodological quality of studies using classical test theory or IRT. In addition, some of the items used by the COSMIN can be subjective, thus, each study was assessed and rated independently by two reviewers to overcome this issue. Again, one reviewer (NK) assessed all studies and trained all assessors prior to data extraction and psychometric assessment to ensure consistency in assessment. All assessments were discussed between the two initial reviewers, and a third was involved if agreement could not be reached.

To compare tools based on an overall psychometric evaluation, the research team developed an ad hoc scoring system. However, this should be regarded as a qualitative rather than quantitative evaluation of the psychometric properties of the scales and allows readers to make broad evaluations about the scales. However, specific scales may have particular merits or drawbacks not revealed by the global score, so readers should look at the evidence for those scales and draw their own conclusions about which scale may be best for their study or clinical practice.

Finally, this review focused on studies that assessed the quality of care and of dying and death retrospectively. Therefore, the psychometric evaluations of the tools identified in this review are based on proxy ratings by family carers and HCPs of how they perceived these experiences following death. Thus, this review is limited to the psychometric evaluation of tools completed after a death and is not applicable to the psychometric properties of these tools assessed before death has occurred.

4.3 Implications

The availability of well-developed and validated tools for quality of care at the end of life and quality of dying and

death are important for several reasons. First, if ‘gold standard’ tools of these constructs existed, comparisons across studies and cultures would enable a better understanding of the similarities and differences between settings. Second, a global measure would eliminate the use of diverse benchmarks for classifying what represents a ‘good’ or ‘bad’ death, for example, it has been previously suggested that dying in the place of preference suggests the person experienced a ‘good’ death [53], but this may not always be the case. Additionally, some of the tools identified in this review are designed for specific populations such as dementia [24] and thus may not be transferable across clinical populations. Finally, tools are essential for evaluating interventions designed to improve quality of care and quality of dying and death. Poorly designed and validated tools may compromise how results of an intervention are interpreted. Therefore, reviews of this kind are important and highlight that, although many tools of these constructs are available, more work remains to be done on validating and improving the psychometric properties of these tools. Conducting research with palliative care populations, whether before or after death, can be uniquely challenging. Researchers and clinicians can use the information provided by this review as a whole or as provided for each measure to make appropriate decisions for which measure would be best suited for their purpose and how similar tools may compare.

5 Conclusion

This systematic review has identified and critically appraised tools for assessing, following death, the quality of care at the end of life and of dying and death. This evaluation has demonstrated that a limited number of tools exists and that they show some promising psychometric properties but still need further investigation. Despite the abundance of tools available to assess the quality of dying and death and satisfaction with care at the end of life, many gaps remain in our understanding of the psychometric properties of these tools. Future research, rather than seeking to develop new tools, might more productively focus on improving and validating existing tools.

Acknowledgements The authors thank Sarah Davis and Jane Harrington and other members of the I-CAN-CARE team, past and present, for their support in completing this research. The authors would like to thank Drs Catriona Mayland and Marco Maltoni for responding to requests for additional information on their studies.

Author Contributions PS and ELS conceived the research, obtained funding for the I-CAN-CARE programme, and managed all elements of the work. PS, NK, BC, and BV contributed to the design of the study. NK and BC developed the search terms. NK completed the database searches, reviewed all titles, abstracts, and full-text studies, extracted data from all studies, managed secondary data extractors,

and synthesised the data. TA, Jane Harrington, and BV each screened a proportion of the titles and abstracts. BC and TA oversaw the full-text review of studies and data extraction. TA oversaw the quality appraisal and data synthesis. TA, GT-R, GS, NW, SH, and TF completed secondary data extraction and quality appraisal tasks. NK drafted the manuscript, and all authors provided critical review on the manuscript and read and approved the final manuscript.

Compliance with Ethical Standards

Conflicts of interest Nuriye Kupeli, Bridget Candy, Gabrielle Tamura-Rose, Guy Schofield, Natalie Webber, Stephanie E Hicks, Theodore Floyd, Bella Vivat, Elizabeth L Sampson, Patrick Stone and Trefor Aspden have no conflicts of interest.

Funding The improving care, assessment, communication and training at the end-of-life (I-CAN-CARE) programme is funded by Marie Curie (grant reference: MCCC-FPO-16-U).

Data Availability Statement All data generated or analysed during this study are included in this published article (and its ESM files).

Open Access This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.


References

1. Etkind S, et al. How many people will need palliative care in 2040? Past trends, future projections and implications for services. *BMC Med.* 2017;15(1):102.
2. Kehl KA, Kowalkowski JA. A systematic review of the prevalence of signs of impending death and symptoms in the last 2 weeks of life. *Am J Hosp Palliat Med.* 2013;30(6):601–16.
3. Hales S, Zimmermann C, Rodin G. Review: the quality of dying and death: a systematic review of measures. *Palliat Med.* 2010;24(2):127–44.
4. Lendon JP, et al. Measuring experience with end-of-life care: a systematic literature review. *J Pain Symptom Manag.* 2015;49(5):904–15 (e3).
5. van Soest-Poortvliet MC, et al. Measuring the quality of dying and quality of care when dying in long-term care settings: a qualitative content analysis of available instruments. *J Pain Symptom Manag.* 2011;42(6):852–63.
6. van Soest-Poortvliet MC, et al. Selecting the best instruments to measure quality of end-of-life care and quality of dying in long term care. *J Am Med Dir Assoc.* 2013;14(3):179–86.
7. Zimmerman S, et al. Measuring end-of-life care and outcomes in residential care/assisted living and nursing homes. *J Pain Symptom Manag.* 2015;49(4):666–79.
8. Mokkink LB, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19(4):539–49.
9. Mokkink LB, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63(7):737–45.

10. De Vet H, et al. *Measurement in medicine: a practical guide*. Cambridge: Cambridge University Press; 2011.
11. Ellis-Smith C, et al. Measures to assess commonly experienced symptoms for people with dementia in long-term care settings: a systematic review. *BMC Med*. 2016;14(1):38.
12. Aspden T, et al. Quality-of-life measures for use within care homes: a systematic review of their measurement properties. *Age Ageing*. 2014;43(5):596–603.
13. Pusic AL, et al. Quality of life among breast cancer patients with lymphedema: a systematic review of patient-reported outcome instruments and outcomes. *J Cancer Surviv*. 2013;7(1):83–92.
14. Albers G, et al. Evaluation of quality-of-life measures for use in palliative care: a systematic review. *Palliat Med*. 2010;24(1):17–37.
15. Moher D, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
16. Candy B et al. *Laxatives for the management of constipation in people receiving palliative care*. The Cochrane Library. 2015.
17. Nekolaichuk CL, et al. Assessing the reliability of patient, nurse, and family caregiver symptom ratings in hospitalized advanced cancer patients. *J Clin Oncol*. 1999;17(11):3621–30.
18. Aminoff BZ, et al. Measuring the suffering of end-stage dementia: reliability and validity of the Mini-Suffering State Examination. *Arch Gerontol Geriatr*. 2004;38(2):123–30.
19. Hearn J, Higginson I. Development and validation of a core outcome measure for palliative care: the palliative care outcome scale. Palliative Care Core Audit Project Advisory Group. *Qual Health Care*. 1999;8(4):219–27.
20. Siegert RJ, et al. Psychological well-being and quality of care: a factor-analytic examination of the palliative care outcome scale. *J Pain Symptom Manage*. 2010;40(1):67–74.
21. Terwee CB, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651–7.
22. Terwee CB, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42.
23. Van Tulder M, et al. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. *Spine*. 2003;28(12):1290–9.
24. Volicer L, Hurley AC, Blasi ZV. Scales for evaluation of end-of-life care in dementia. *Alzheimer Dis Assoc Disord*. 2001;15(4):194–200.
25. Vohra JU, et al. Family perceptions of end-of-life care in long-term care facilities. *J Palliat Care*. 2004;20(4):297.
26. Teno JM, et al. Validation of toolkit after-death bereaved family member interview. *J Pain Symptom Manage*. 2001;22(3):752–8.
27. Mayland CR, et al. Assessing the quality of care for dying patients from the bereaved relatives' perspective: further validation of "evaluating care and health outcomes—for the dying". *J Pain Symptom Manage*. 2014;47(4):687–96.
28. Mayland CR, Williams EM, Ellershaw JE. Assessing quality of care for the dying: the development and initial validation of a postal self-completion questionnaire for bereaved relatives. *Palliat Med*. 2012;26(7):897–907.
29. Casarett D, et al. A nationwide VA palliative care quality measure: the family assessment of treatment at the end of life. *J Palliat Med*. 2008;11(1):68–75.
30. Ornstein KA, et al. Use of an item bank to develop two short-form FAMCARE scales to measure family satisfaction with care in the setting of serious illness. *Journal of pain and symptom management*. 2015;49(5):894–903 (e4).
31. Mayland CR, et al. Caring for those who die at home: the use and validation of 'Care Of the Dying Evaluation'(CODE) with bereaved relatives. *BMJ Support Palliat Care*. 2014;4(2):167–74.
32. Miyashita M, et al. Development the Care Evaluation Scale Version 2.0: a modified version of a measure for bereaved family members to evaluate the structure and process of palliative care for cancer patient. *BMC Palliat Care*. 2017;16(1):8.
33. Claessen SJ, et al. Measuring Relatives' perspectives on the quality of palliative care: the consumer quality index palliative care. *J Pain Symptom Manage*. 2013;45(5):875–84.
34. Partinico M, et al. A new Italian questionnaire to assess caregivers of cancer patients' satisfaction with palliative care: multicenter validation of the post mortem questionnaire-short form. *J Pain Symptom Manage*. 2014;47(2):298–306.
35. Morita T, Chihara S, Kashiwagi T. A scale to measure satisfaction of bereaved family receiving inpatient palliative care. *Palliat Med*. 2002;16(2):141–50.
36. Morita T, et al. Measuring the quality of structure and process in end-of-life care from the bereaved family perspective. *J Pain Symptom Manage*. 2004;27(6):492–501.
37. Ringdal GI, Jordhøy MS, Kaasa S. Measuring quality of palliative care: psychometric properties of the FAMCARE Scale. *Qual Life Res*. 2003;12(2):167–76.
38. Casarett D, et al. Measuring families' perceptions of care across a health care system: preliminary experience with the Family Assessment of Treatment at End of Life Short form (FATE-S). *J Pain Symptom Manage*. 2010;40(6):801–9.
39. van Soest-Poortvliet MC, 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):671–84.
40. Shin DW, et al. Measuring the structure and process of end-of-life care in Korea: validation of the Korean version of the Care Evaluation Scale (CES). *J Pain Symptom Manage*. 2012;44(4):615–25 (e2).
41. Higgins PC, Prigerson HG. Caregiver evaluation of the quality of end-of-life care (CEQUEL) scale: the caregiver's perception of patient care near death. *PLoS One*. 2013;8(6):e66066.
42. Curtis JR, et al. A measure of the quality of dying and death: initial validation using after-death interviews with family members. *J Pain Symptom Manage*. 2002;24(1):17–31.
43. Downey L, et al. The Quality of Dying and Death Questionnaire (QODD): empirical domains and theoretical perspectives. *J Pain Symptom Manage*. 2010;39(1):9–22.
44. Heckel M, et al. Validation of the German Version of the Quality of Dying and Death Questionnaire for Informal Caregivers (QODD-D-Ang). *J Pain Symptom Manage*. 2015;50(3):402–13.
45. Heckel M, et al. Validation of the German version of the quality of dying and death questionnaire for health professionals. *Am J Hosp Palliat Med*. 2016;33(8):760–9.
46. Pérez-Cruz PE, et al. Validation of the Spanish Version of the Quality of Dying and Death Questionnaire (QODD-ESP) in a home-based cancer palliative care program and development of the QODD-ESP-12. *J Pain Symptom Manage*. 2017;53:1042–9.
47. Cornally N, et al. Measuring staff perception of end-of-life experience of older adults in long-term care. *Appl Nurs Res*. 2016;30:245–51.
48. Munn JC, et al. Measuring the quality of dying in long-term care. *J Am Geriatr Soc*. 2007;55(9):1371–9.
49. Shin DW, et al. Measuring comprehensive outcomes in palliative care: validation of the Korean version of the Good Death Inventory. *J Pain Symptom Manage*. 2011;42(4):632–42.
50. Cagle JG, et al. Validation of the quality of dying-hospice scale. *J Pain Symptom Manage*. 2015;49(2):265–76.
51. Miyashita M, et al. Good death inventory: a measure for evaluating good death from the bereaved family member's perspective. *J Pain Symptom Manage*. 2008;35(5):486–98.
52. Terwee CB, et al. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res*. 2016;25(4):767–79.

53. McWhinney IR, Bass MJ, Orr V. Factors associated with location of death (home or hospital) of patients referred to a palliative care team. *CMAJ*. 1995;152(3):361.
54. McCusker J. Development of scales to measure satisfaction and preferences regarding long-term and terminal care. *Med Care*. 1984;22(5):476–93.
55. Kiely DK, et al. The validity and reliability of scales for the evaluation of end-of-life care in advanced dementia. *Alzheimer Dis Assoc Disord*. 2006;20(3):176.
56. Van Der Steen J, et al. Ratings of symptoms and comfort in dementia patients at the end of life: comparison of nurses and families. *Palliat Med*. 2009;23(4):317–24.
57. 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(1):565–74.

Affiliations

Nuriye Kupeli¹  · **Bridget Candy¹** · **Gabrielle Tamura-Rose²** · **Guy Schofield³** · **Natalie Webber⁴** · **Stephanie E. Hicks⁵** · **Theodore Floyd⁶** · **Bella Vivat¹** · **Elizabeth L. Sampson^{1,7}** · **Patrick Stone¹** · **Trefor Aspden⁸**

¹ Division of Psychiatry, Marie Curie Palliative Care Research Department, University College London, 6th Floor, Maple House, 149 Tottenham Court Road, London W1T 7NF, UK

² Department of Palliative Care, Royal Free Hospital, London, UK

³ Centre for Ethics in Medicine, University of Bristol, Bristol, UK

⁴ Royal Trinity Hospice, London, UK

⁵ St Christopher's Hospice, London, UK

⁶ Emergency Department, Medway Maritime Hospital NHS Trust, Kent, UK

⁷ Barnet Enfield and Haringey Mental Health Trust Liaison Psychiatry Team, North Middlesex University Hospital, London, UK

⁸ Division of Psychiatry, University College London, London, UK