Psychometric properties of instruments to measure the quality of end-of-life care and dying for long-term care residents with dementia

**Mirjam C. van Soest-Poortvliet · Jenny T. van der Steen · Sheryl Zimmerman ·
Lauren W. Cohen · Maartje S. Klapwijk · Mirjam Bezemer · Wilco P. Achterberg ·
Dirk L. Knol · Miel W. Ribbe · Henrica C. W. de Vet**

Abstract

Purpose Quality of care for long-term care (LTC) residents with dementia at the end-of-life is often evaluated using standardized instruments that were not developed for or thoroughly tested in this population. Given the importance of using appropriate instruments to evaluate the quality of care (QOC) and quality of dying (QOD) in LTC, we compared the validity and reliability of ten available instruments commonly used for these purposes.

Methods We performed prospective observations and retrospective interviews and surveys of family ($n = 70$)
and professionals ($n = 103$) of LTC decedents with dementia in the Netherlands.

Results Instruments within the constructs QOC and QOD were highly correlated, and showed moderate to high correlation with overall assessments of QOC and QOD. Prospective and retrospective ratings using the same instruments differed little. Concordance between family and professional scores was low. Cronbach's alpha was mostly adequate. The EOLD–CAD showed good fit with pre-assumed factor structures. The EOLD–SWC and FPCS appear most valid and reliable for measuring QOC, and the EOLD–CAD and MSSE for measuring QOD. The POS performed worst in this population.

Conclusions Our comparative study of psychometric properties of instruments allows for informed selection of QOC and QOD measures for LTC residents with dementia.

M. C. van Soest-Poortvliet (✉) · J. T. van der Steen ·
M. W. Ribbe
Department of Nursing Home Medicine, EMGO Institute
for Health and Care Research, Amsterdam, The Netherlands
e-mail: m.vansoest@vumc.nl

J. T. van der Steen
Department of Public and Occupational Health, EMGO Institute
for Health and Care Research, Amsterdam, The Netherlands

S. Zimmerman · L. W. Cohen
Cecil G. Sheps Center for Health Services Research,
University of North Carolina, Chapel Hill, NC, USA

M. S. Klapwijk
Nursing Home Van Wijckerslooth, Oegstgeest, The Netherlands

M. Bezemer
Zorgspectrum, Nieuwegein, Houten, The Netherlands

W. P. Achterberg
Department of Public Health and Primary Care,
Leiden University Medical Center, Leiden, The Netherlands

D. L. Knol · H. C. W. de Vet
Department of Epidemiology and Biostatistics, EMGO Institute
for Health and Care Research, Amsterdam, The Netherlands

Abbreviations

AL	Assisted living
BANS–S	Bedford alzheimer nursing severity–scale
CCC	Concordance correlation coefficients
CFA	Confirmatory factor analysis
CFI	Comparative fit index
DS–DAT	Discomfort scale–dementia of alzheimer type
DSI	Decision satisfaction inventory
EOLD	End-of-life in dementia
EOLD–CAD	EOLD–comfort assessment in dying
EOLD–SM	EOLD–symptom management
EOLD–SWC	EOLD–satisfaction with care
FATE–S	Family assessment of treatment at the end-of-life–short version

FPCS	Family perceptions of care scale
FPPFC	Family perception of physician-family caregiver communication
LTC	Long-term care
MDS–PC	Minimum data–set palliative care
MSSE	Mini-suffering state examination
NH	Nursing home
PAINAD	Pain assessment in advanced dementia
POS	Palliative care outcome scale
QOC	Quality of care
QOD	Quality of dying
QOD–LTC	Quality of dying in long-term care
RC	Residential care home
RMSEA	Root mean square error of approximation
SD	Standard deviation
STAS	Support team assessment schedule
TLI	Tucker–Lewis index
TIME	Toolkit of instruments to measure end-of-life care

Introduction

Many decedents in long-term care (LTC) settings have dementia, with 50–92% of individuals with dementia dying in nursing homes (NHs) across several countries [1, 2]. Unfortunately, numerous shortcomings exist in end-of-life care, including resident suffering and unmet family needs [3, 4]. To identify and address these shortcomings in clinical practice and research, measurement instruments specific for LTC settings are needed. These instruments should be useful for residents with dementia because of the high number of people dying with dementia in these settings and the projected increase of this population [5, 6]. Further, this population is especially at risk for poor care due to challenges of communication and assessment. For example, pain and other symptoms may not be expressed verbally, and so must instead be detected by more subtle behavioral cues.

When evaluating quality at the end-of-life of people with dementia in LTC, two constructs, "quality of care (QOC)" and "quality of dying (QOD)" must be distinguished. QOC reflects elements of the setting in which dying takes place, such as the availability of professionals, types of communication and scope of treatment. QOD refers to symptom burden and other resident experiences potentially influenced by care, and is also partly the result of various resident-related factors. For example, an individual's QOD may be affected by pain (symptom burden) that was treated with medications and massage (care), as well by his cognitive status (patient factor). QOD may be

considered synonymous with quality of life while dying, and while the time period to which this refers cannot be precisely defined, measurement tools often encapsulate the period to encompass a specific period of time prior to death. Both, QOC and QOD instruments, could be used to evaluate quality at the end-of-life toward the development of quality indicators.

Ten instruments are available to measure quality when dying with dementia in LTC and which have been developed or used in populations with a substantial number of dementia residents [7]. As these measures have become available only after 2000, little is known about their psychometric properties when used with this group. Table 1 presents characteristics and previously published psychometric properties of these ten measurements, including findings from samples of residents without dementia. This paper extends our knowledge of these instruments by assessing and comparing the validity and reliability of the measures of perceived QOC and QOD for residents dying with dementia in NHs and residential care homes (RCs).

Methods

The tested instruments

Instruments that assess QOC

The end-of-life in dementia satisfaction with care (EOLD–SWC) was developed for after-death assessment of satisfaction with care by family of dementia patients [8]. Higher scores reflect higher satisfaction. The family assessment of treatment at the end-of-life–short version (FATE–S) was developed to evaluate respondents' perceptions of outcomes of care, asking how well the provided care met resident or family needs. FATE–S scores are expressed as a percentage of valid responses for which families provided the best possible response; higher percentages reflect better treatment [11]. The family perception of care scale (FPCS) asks the family's opinion about the quality of resident care, family support, communication, and rooming [13]. Higher summed total scores reflect more favorable perceptions of end-of-life care. The family perception of physician–family caregiver communication (FPPFC) was designed to assess family perception of physician communication during the dying process, and higher scores reflect higher quality of family–physician communication [14]. The nursing home version of the after-death bereaved family member interview of the toolkit of instruments to measure end-of-life care (TIME) consists of items asking family about seven domains of care [15]. For five of the domains, the questions are summarized as "problem scores," with higher numbers reflecting more opportunity to improve. For the physical

Table 1 Characteristics and psychometric properties of ten available measurement instruments reported in the literature

Measurement instrument (range)	Subscales	Items	Response options	Population in development	Relation to other tools	Internal consistency	Factor analysis	Respondent type in this (and original) study	Recall period in this (and original) study
QOC									
EOLD–SWC (10–40) [8–10]	No	10	Range 1–4	NH residents with advanced dementia	$r = 0.80$ with DSI	$\alpha = 0.83$	One factor only	Family (originally: family)	Last month (originally: 3 months)
FATE–S (0–100) [11, 12]	No	12	Range 0–100	Eligible patients who received care from a veteran facility	$r = 0.89$ with FATE	No results available for the FATE-S FATE: $\alpha = 0.91$	No results available	Family (originally: family)	Last month (originally: last month)
FPCS (25–175) [13]	Resident care, family support, communication, and rooming	25	Range 1–7	Residents of LTC facilities		$\alpha = 0.96$	Four factors; $r = -0.66–0.53$	Family (originally: family)	Last month (originally: last month)
FPPFC (1–4) [14]	No	7	Range 1–4	Family of NH and RC/AL residents		$\alpha = 0.96$	No results available	Family (originally: family)	Last 3 months (originally: last 3 months)
TIME (varies per domain)[a] [15, 16]	Physical comfort and emotional support, inform and promote shared decision making, encourage advance care planning, focus on individual, attend to the emotional and spiritual needs of the family, provide coordination of care, support the self-efficacy of the family, and overall rating scale for patient focused, family centered care	36	Various	Patients of hospice, hospital and NH	$r = 0.23–0.59$ with single rating item for each domain	$\alpha = 0.58–0.87$ with eight domains	No results available	Family (originally: family)	Last week (originally: last week)
QOD									
EOLD–CAD (14–42) [9, 10]	Physical distress, dying symptoms, emotional symptoms and well-being	14	Range 1–3	NH residents with advanced dementia	$r = -0.5$ to -0.8 with MSSE [17]	$\alpha = 0.82$	Four factors; $r = 0.71–0.82$	Family/Professionals (originally: family)	Last week (originally: dying)
EOLD–SM (0–45) [9, 10]	No	9	Range 0–5	NH residents with advanced dementia	No results available	$\alpha = 0.68$	Two factors; $r = 0.47–0.81$	Family/Professionals (originally: family)	Last month (originally: 3 months)

Table 1 continued

Measurement instrument (range)	Subscales	Items	Response options	Population in development	Relation to other tools	Internal consistency	Factor analysis	Respondent type in this (and original) study	Recall period in this (and original) study
MSSE (0–10/11)[a] [17]	No	10/11	Yes or no	Patients with dementia of the Stuchynski Alzheimer Research and Treatment Center	$r = -0.5$ to -0.8 with CAD	$\alpha = 0.70$	No results available	Family/Professionals (originally: physicians)	Last week (originally: no recall period; prospective assessment)
POS (0–4)[a] [18, 19]	No	10	Range 0–4	Patients of centers providing palliative care	$r = 0.43$–0.53 with comprehensive quality of life measurement instrument, $r = 0.51$–0.80 with STAS	$\alpha = 0.65$ patient version $\alpha = 0.70$ professional caregivers report	Two factors and 3 single items; $r = 0.56$–0.80 [20]	Family/Professionals (originally: patient, family, professionals)	Last 3 days (originally: past 3 days; prospective assessment)
QOC and QOD									
QOD-LTC (1–5) [21]	Personhood, closure and preparatory tasks	11	Range 1–5	Residents of RC/AL settings and NHs; 73% cognitively impaired	No results available	$\alpha = 0.66$ full scale $\alpha = 0.50$–0.57 subscales	Three factors; $r = 0.38$–0.77	Family/Professionals (originally: family and professionals)	Last week (originally: last week)

[a] For some TIME domain scores, the MSSE and POS, a lower score reflects higher quality. For all other instruments a higher score reflects higher quality

DSI Decision satisfaction inventory, *STAS* support team assessment schedule, *LTC* long-term care, *RC* resident care, *AL* assisted living, *NH* nursing home, *EOLD* end-of-life in dementia, *EOLD–CAD* EOLD-comfort assessment in dying, *EOLD–SM* EOLD-symptom management, *EOLD–SWC* EOLD-satisfaction with care, *FATE–S* family assessment of treatment at the end-of-life short version, *FPCS* family perceptions of care scale, *FPPFC* family perception of physician–family caregiver communication, *MSSE* mini-suffering state examination, *POS* palliative care outcome scale, *QOD–LTC* quality of dying in long-term care, *TIME* nursing home version of the Toolkit of instruments to Measure End-of-Life Care after-death bereaved family interview

comfort and emotional support domain, questions are not summarized. For the support self-efficacy of the family domain, questions are summarized on a 3-point scale, with higher scores reflecting more support. Additionally, the instrument produces an overall rating scale for patient-focused, family-centered care, with a higher score being favorable. Cut points are not available for any of the QOC instruments.

Instruments that assess QOD

The EOLD–comfort assessment at death (EOLD–CAD) comprises the subscales physical distress, dying symptoms, emotional symptoms, and well-being; the EOLD–symptom management (EOLD–SM) assesses the extent to which symptoms were experienced [8]. Although no cut-off scores are available for either instrument, higher scores reflect more comfort and better symptom management, respectively. Both were originally developed for dementia patients. The mini-suffering state examination (MSSE) was developed as a prospective assessment to reflect the amount of suffering experienced by dementia patients, with higher scores indicating more suffering [17], but can also be used retrospectively. The developers considered scores of 0–3 as a low level of suffering; 4–6 as intermediate; and 7–10 as a high level of suffering. The palliative care outcome scale (POS) was developed for cancer patients, but has also been used for cognitively impaired patients [19]. It concerns the physical, psychological, and spiritual domains of life within the field of palliative care [18, 19]. The POS developers considered a mean score of 2.0 or lower as favorable.

Instruments that assess the QOC and QOD

The quality of dying–long-term care (QOD–LTC) was developed for cognitively impaired and intact residents in NHs and residential care/assisted living settings [21]. It assesses perspectives on quality of personhood, closure, and preparatory tasks. Higher mean scores reflect a higher quality of end-of-life in LTC. A cut-off score is not available.

Translation

For the EOLD-instruments and MSSE, valid Dutch versions were available [22, 23]. A Dutch version of the POS was also available [19], but to adhere to standard norms, the translation process was repeated. Two researchers (MvS-P and JTS) independently translated the English versions of the POS and other instruments into Dutch in consultation with the developers when necessary; they resolved differences in translations by reviewing and

discussing each differently translated word, iteratively, until agreement was reached. The instruments were back translated by a professional translator.

Data collection

The study was conducted between March 2008 and April 2009 in seven NHs and seven RCs in the Netherlands. Eligible resident cases were those who died with dementia within this time period, excluding a single case who stayed in the hospital for more than half of the last month of life.

For each eligible resident, retrospective interviews were completed with a professional (within 2 weeks of death) and a family caregiver (approximately 2 months after death). Two weeks assured that professional caregivers would remember that particular patient and have access to records, while 2 months allowed the family time to grieve yet was not too long to affect recall. These measurement time points were the same as for the CASCADE study [24], and used in a recent Dutch study [4]. The professional caregiver interview included the EOLD–CAD, EOLD–SM, MSSE, POS, and QOD–LTC, as well as the seven-item Bedford Alzheimer nursing severity–scale (BANS–S) [25], which assessed dementia status 1 month pre-death. BANS–S scores (range 7–28) of 17 and higher are regarded as severe dementia [26]. In NHs, specialty physicians (referred to as elderly care physicians) completed the questionnaires, and in four of the seven NHs, a nurse completed it independently of the physician. In RC, elderly care physicians completed the questionnaire in consultation with nurses. According to the Dutch model of care, elderly care physicians are on-staff in NHs, and they are also responsible for dementia care units in RCs [27].

Family received an informed consent form by postal mail approximately 6 weeks after death of their family member. When they agreed to participate, they received a mailed questionnaire with the EOLD–SWC, EOLD–SM, EOLD–CAD, FPPFC, MSSE, POS, and QOD–LTC. The MSSE contains 10 items, one of which is the clinician's opinion about suffering. We adapted the instrument for families, separately asking the physician's and nurse's opinion. The instruments were compiled in order of time-frame of reference, beginning with instruments referring to the last month of life, and ending with instruments focusing on the last 3 days of life. To avoid possible bias due to the order of instruments within each timeframe, for half of the population the order within timeframes was reversed. Four overall assessments of family perceptions preceded the other measures: "On a scale of 1–10, where 1 = the worst possible and 10 = the best possible, how would you rate the overall quality of care (quality of life) in the last month (last days) of life?" Additionally, family provided resident demographic data. After the family completed the written

survey, they were contacted by telephone to complete the FATE–S, FPCS and TIME. The TIME was administered by phone due to its development as a telephone interview and its complexity; the FATE–S and FPCS were identified midway through the study and so were included only in the last 23 family interviews.

In addition to retrospective questionnaires and interviews, in two NHs, 24 residents who were expected to die were observed prospectively by the coordinating elderly care physicians twice daily when possible; these data were regarded as the gold standard and used to assess differences in retrospective and prospective reporting. The physicians completed the MSSE, the EOLD–CAD, the discomfort scale–dementia of Alzheimer type (DS–DAT), [28] and the pain assessment in advanced dementia (PAINAD) scale [29]. These measures specifically assess discomfort and pain in dementia. The coordinating elderly care physicians were trained in use of the DS–DAT and PAINAD with an instructional video, and during the training session, practiced scoring of observational instruments referring to a gold standard. The DS–DAT has four response option (range 0–3) indicating frequency and intensity of behavior; nine items are summed to calculate a total score (range 0–27) with a higher score reflecting more discomfort. The PAINAD has three response options (range 0–2); five items are summed to calculate a total score (range 0–10) with a higher score reflecting more pain. Mean scores were calculated for all of these instruments averaging the two assessments per day and over days.

Data analysis

Validity

Validity was assessed by: (1) comparing prospective observed discomfort and symptoms with prospective and retrospective measurements of the EOLD–CAD and MSSE (concurrent validity); (2) comparing the average (sub)scale scores on the QOC instruments and the overall assessment of QOC and similarly comparing the QOD instruments to the overall assessment of QOD (convergent validity); (3) comparing total scores of instruments measuring similar constructs (QOC or QOD), and total scores of the QOD instruments with the total scores of DS–DAT and PAINAD (convergent validity); and (4) confirmatory factor analysis (CFA) to assess the factor structure of the instruments (factorial validity).

First, we compared total scores on the EOLD–CAD and MSSE administered prospectively before death with total scores on the same instruments retrospectively after death in the same residents. Referring to prospective measurement as the gold standard, we assumed EOLD–CAD and MSSE total scores to be comparable to those assessed in retrospect. To examine the agreement between prospective observations

and retrospective ratings by professionals and family, we expressed the differences in effect sizes (d) and interpreted these according to Cohen (1988): small effect when d is between 0.2 and 0.3, moderate effect when d is around 0.5 and large effect when $d \geq 0.8$ [30]. Additionally, concordance correlation coefficients (CCC) were calculated for total scores, including 95% bias corrected accelerated (BCa) bootstrap confidence intervals (CI) [31, 32]. A CCC of <0.00 represents poor concordance; 0.00–0.20 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; and 0.81–1.00 almost perfect concordance [33].

Second, the correlation between the average (sub)scale scores on the QOC instruments and the overall QOC assessment was calculated to determine whether the instruments measure the expected construct (high correlation in the expected direction). Similarly, we calculated correlations between the QOD instruments and the overall assessment. We selected the appropriate construct and closest timeframe (1 month or last days for the overall assessments). For instruments about the last week of life, we calculated the correlation with the overall assessment of the last month as well as the last days. Correlation coefficients ≥0.5 represent high correlations, 0.3–0.5 moderate, 0.3–0.1 small, and <0.1 insubstantial correlations [30].

Third, we examined the intercorrelation of total scores of instruments measuring similar constructs, as well as the total scores of QOD instruments with DS–DAT and PAINAD total scores.

Lastly, CFA was conducted for the QOD instruments completed by professionals to confirm the structural model as reported by the instruments' developers. No CFA was performed for the family data due to the small sample size (CFA requires a minimum of 100 respondents). Further, CFA is only relevant when the instrument is based (or likely based) on a reflective model, meaning that the items are reflections of the same construct [34, 35]. For example, we did not assess CFA on the MSSE because this instrument comprises a collection of different items that cause suffering in a resident which are not expected to correlate. Similarly, we did not apply CFA to the POS. We examined whether predetermined scales by the developers (four for the EOLD–CAD, and three for the QOD–LTC) could be confirmed in our analyses. Tucker–Lewis index (TLI)/comparative fit index (CFI) values >0.95 represent good fit [36]. A root mean square error of approximation (RMSEA) value of <0.08 represents acceptable fit and <0.05 represents good fit [36]. For the RMSEA, 90% CIs were also calculated.

Reliability

Internal consistency was assessed for family and professionals separately, calculating Cronbach's alpha for all instruments based (or likely based) on a reflective model.
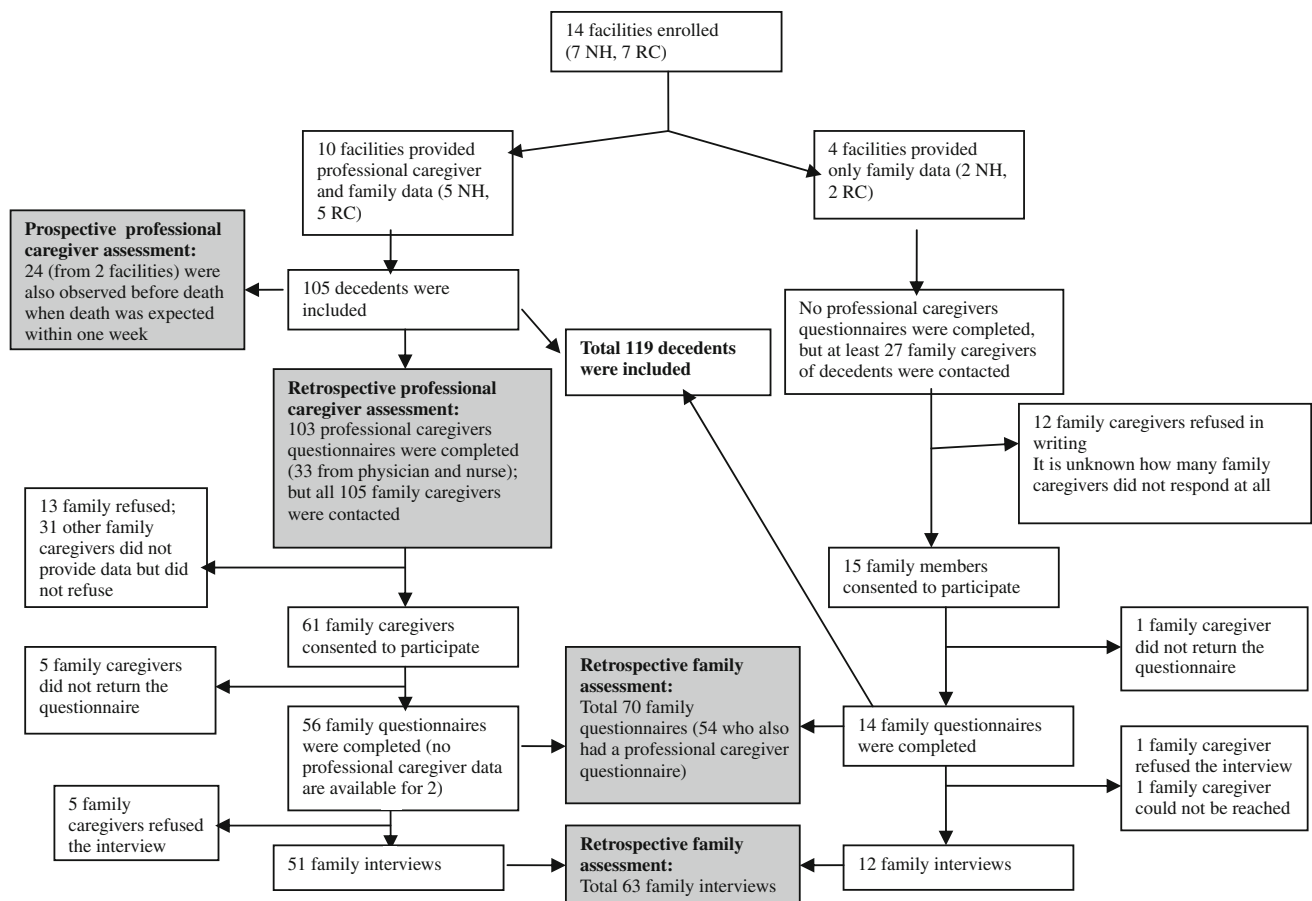
Fig. 1 Overview of the data collection to evaluate properties of instruments

Table 2 Prospective observation by elderly care physicians versus retrospective completion of the same QOD instruments, by respondent type

Measure (range)	N	Mean scores (SD)[a]		P value mean diff	d	CCC (95% CI)
		Prospective observations	Retrospective questionnaire			
		Physician	Physician			
EOLD–CAD (14–42)	13	39.1 (3.1)	36.0 (3.1)	0.16	1.00	0.20 (−0.18; 0.61)
MSSE (0–8[b])	13	2.0 (1.4)	2.5 (1.8)	0.32	0.31	0.68 (0.20; 0.90)
		Physician	Nurse			
EOLD–CAD (14–42)	16	38.6 (4.7)	34.3 (5.9)	0.03	0.81	0.44 (−0.05; 0.77)
MSSE (0–8[b])	16	1.8 (1.2)	3.2 (1.6)	0.12	0.99	0.28 (−0.14; 0.68)
		Physician	Family			
EOLD–CAD (14–42)	15	39.0 (5.0)	31.0 (5.3)	0.00	1.55	0.27 (−0.10; 0.68)
MSSE (0–8[b])	15	1.4 (1.0)	3.1 (2.0)	0.05	1.08	0.25 (−0.14; 0.65)

CCC Concordance correlation coefficient, *CI* confidence interval, *d* effect size, *EOLD–CAD* end-of-life in dementia-comfort assessment in dying, *MSSE* mini-suffering state examination, *SD* standard deviation

[a] For the EOLD–CAD, a higher score signifies a higher quality and for the MSSE, a lower score reflects a higher quality

[b] The opinions about suffering from different viewpoints were not included in this comparison

We computed alphas for the FPPFC, the subscales of the EOLD–CAD, FPCS and QOD–LTC, and the TIME domains. Alphas between 0.70 and 0.95 were considered adequate [37].

Additionally, differences in ratings of family, nurses and physicians were tested with paired sample t tests when both professional and family caregiver data were available. To assess the professional and family caregiver, concordance

Table 3 Correlation of instrument scores completed retrospectively by family caregiver with overall assessments of the QOC/QOD ($N = 70$)

	Pearson correlations with overall assessment			
	QOC		QOD	
	Last month	Last days	Last month	Last days
QOC				
EOLD–SWC	0.70	–	–	–
FATE–S	0.49			
FPCS	0.65	–	–	–
Resident care	0.68	–	–	–
Family support	0.37	–	–	–
Communication	0.71	–	–	–
Rooming	0.26	–	–	–
FPPFC[a]	0.40	–	–	–
TIME[b,c]	−0.44 to 0.78	−0.46 to 0.67	–	–
QOD				
EOLD–CAD	–	–	0.33	0.27
Physical distress	–	–	0.47	0.30
Dying symptoms	–	–	0.25	0.22
Emotional distress	–	–	0.13	0.20
Well-being	–	–	−0.03	−0.10
EOLD–SM	–	–	0.36	–
MSSE[c]	–	–	−0.38	−0.36
POS[c]	–	–	–	−0.38
QOC and QOD				
QOD–LTC[d]	–	–	–	–
Personhood	0.69	–	–	–
Closure	–	–	0.36	–
Preparatory tasks	0.08	–	–	–

Numbers were lower for instruments assessed by interview: the TIME ($n = 63$), FPCS ($n = 23$ last decedents) and FATE-S ($N = 22$ last decedents)

Abbreviations of instruments are listed under Table 1

[a] FPPFC asks about the last 3 months of life

[b] Range of correlation coefficients with TIME domain scores, because there is no total score available for the TIME

[c] For some TIME domain scores, the MSSE and POS, a lower score reflects higher quality. For all the other instruments, a higher score reflects higher quality

[d] The subscales personhood and preparatory tasks measure QOC and the subscale closure refers to QOD

of scores in individual residents CCCs were calculated [31, 32].

Imputations and software

When at least 75% of the items were answered, missing values, except those in the FATE–S, were imputed with the subject's mean score. A two-way imputation (with subject mean, item mean, and a random factor), was also conducted, but because it yielded similar results, the simpler subject means were used. CFA was conducted using Mplus version 6.1 [38]. We used SPSS 15.0 for all other analyses (SPSS Inc., Chicago Illinois, USA).

Results

The study included 119 decedents on whose behalf 70 family and 103 professionals (physician and/or nurse) completed an evaluation (Fig. 1). Decedents were predominantly female (82%), widowed (69%), native (96%), with a mean age at death of 88.1 (standard deviation [SD]

Table 4 Correlation of instrument scores completed retrospectively

#	Instrument	Correlations									
		1	2	3	4	5	6	7	8	9	10
a. Correlation matrix of instrument scores provided by family caregivers (N = 70[a])											
QOC											
1	EOLD–SWC	–	–	–	–	–					
2	FATE–S	0.67	–	–	–	–					
3	FPCS	0.61	0.62	–	–	–					
4	FPPFC	0.52	0.39	0.66	–	–					
5	TIME[b,c]	−0.54 to 0.67	−0.45 to 0.54	−0.88 to 0.72	−0.39 to 0.53	–					
QOD											
6	EOLD–CAD						–	–	–	–	–
7	EOLD–SM						0.72	–	–	–	–
8	MSSE[c]						−0.58	−0.53	–	–	–
9	POS[c]						−0.59	−0.38	0.58	–	–
QOC and QOD											
10	QOD–LTC	0.43	0.51	0.56	0.53	−0.52 to 0.46	0.31	−0.30	−0.36	0.49	–

#	Instrument	Correlations				
		1	2	3	4	5
b. Correlation matrix of instrument scores provided by professional caregivers (N = 103)						
QOD						
1	EOLD–CAD	–	–	–	–	–
2	EOLD–SM	0.68	–	–	–	–
3	MSSE[d]	−0.70	−0.67	–	–	–
4	POS[d]	−0.43	0.56	−0.49	–	–
QOC and QOD						
5	QOD–LTC	0.36	−0.38	−0.32	0.41	–

Abbreviations of instruments are listed under Table 1

[a] Numbers were lower for instruments assessed by interview: the TIME (n = 63), FPCS (n = 23 last decedents) and FATE-S (N = 22 last decedents)

[b] Range of correlation coefficients with TIME domain scores, no total score available for the TIME

[c] For some TIME domain scores, the MSSE and POS, a lower score reflects higher quality. For all the other instruments, a higher score reflects higher quality

[d] For the MSSE and POS, a lower score reflects higher quality. For all the other instruments, a higher score reflects higher quality

6.6). Sixty-three percent had severe dementia 1 month pre-death. Sixty-eight percent died in a NH, and 32% in a RC; none died in a hospital. Mean length of stay was 31 months (SD 29 months).

Most family caregivers were female (67%), married (67%), native (99%), and children of the decedent (76%). Their mean age was 60.6 (SD 8.5). In the last week before death, the mean number of visits was 5.2 (SD 2.2), and family caregivers spoke to staff 3.0 times (SD 2.2). In the last month, family caregivers visited the residents 16.8 times (SD 8.6) and spoke to staff 13.5 times (SD 7.9). Of the family caregivers, 39% were present at death. A physician was present in 3% of cases, a nurse in 31%, and another family caregiver was present in 41% of cases. No one was present at death for 27% of the decedents. Death was expected by family caregivers in 39% of cases, whereas professional caregivers expected death in 71% of cases.

Validity

The mean total EOLD–CAD scores (Table 2) and subscale scores (not presented) were not significantly different between prospective and retrospective ratings of physicians, but the retrospective rating of the nurse was significantly different from the prospective physicians' ratings. For the mean total MSSE scores, there were no significant differences. However, prospective observations tended to

Table 5 Correlation of prospective observation with the DS–DAT and PAINAD by elderly care physicians versus instrument scores completed retrospectively by professional caregivers (N = 24 decedents)

Retrospective questionnaire	Prospective observations	
	DS–DAT	PAINAD
Reference instruments		
DS–DAT	0.63	0.58
PAINAD	0.40	0.40
Instruments under study		
EOLD–CAD	−0.25	−0.20
EOLD–SM	−0.13	−0.15
MSSE	0.09	0.14
POS	0.08	0.03

DS–DAT Discomfort scale–dementia of alzheimer type, *PAINAD* pain assessment in advanced dementia, higher scores represent more discomfort or pain; Abbreviations and interpretations of scores of the other instruments as high or low are listed under Table 1

be more favorable for all pairs, and were statistical significantly more favorable when compared with family ratings. The CCC of the MMSE by prospective observation and the physicians' retrospective rating of the MSSE was substantial (CCC = 0.68) and the effect size was moderate ($d = 0.31$). For all other pairs, the concordance of scores in individual residents was slight to moderate (CCC = 0.20–0.44) and the effect size was large ($d = 0.81$–1.55).

Table 3 shows that correlations with the overall assessment were generally higher for the QOC instruments ($r = 0.13$–0.70) than for the QOD instruments ($r = 0.03$–0.47). High correlations with the overall assessment were found for the EOLD–SWC, FPCS (totalscale, subscale resident care and communication), the TIME overall rating scale, and the QOD–LTC subscale personhood. We found a small correlation for the EOLD–CAD subscale emotional distress; insubstantial correlations for the EOLD–CAD well-being subscale and the QOD–LTC subscale preparatory tasks, and the others were moderate. The only direction that was not consistent with our hypothesis was the correlation between the EOLD–CAD scale well-being and the QOD overall assessment. We found mostly moderate to high intercorrelation between instruments measuring the same construct (QOC $r = 0.39$–0.88; QOD $r = 0.38$–0.72), except for some TIME domains (Table 4). The instrument that measures two constructs, the QOD–LTC, correlated somewhat better with QOC instruments than with QOD instruments. Correlations of the QOD instruments with the prospective observations of DS–DAT and PAINAD were low for the EOLD–CAD ($r = 0.20$–0.25), and even more so for the EOLD–SM, MSSE, and POS ($r = 0.03$–0.15) (Table 5).

CFA showed that the fit index (according to CFI and TLI) was good for the EOLD–CAD (CFI and TLI = 0.98), but not for the QOD–LTC (CFI = 0.87 and TFI = 0.83) (Table 6). The EOLD–CAD also showed acceptable fit according to the RMSEA criterion (0.08), but the CIs overlap both good and poor fit (0.05–0.11).

Internal consistency

Cronbach's alpha's for the QOC instruments were adequate ($\alpha = 0.79$–0.91), except for the subscale Rooming of the FPCS ($\alpha = 0.31$), and some domains of the TIME ($\alpha = 0.31$–0.58) (Table 6). For the QOD instruments, the Cronbach's alpha was adequate ($\alpha = 0.70$–0.83), except for the physical distress subscale of the EOLD–CAD ($\alpha = 0.64$). For the QOD–LTC, only the Cronbach's alpha of the Personhood subscale was adequate ($\alpha = 0.70$).

Inter respondent reliability

Mean QOD–LTC, total EOLD–CAD, and EOLD–SM scores were higher (better) for professionals compared with family (Table 7). The effect size of the difference was small for the EOLD–CAD ($d = 0.39$) and moderate for the QOD–LTC ($d = 0.57$) and EOLD–SM ($d = 0.75$). The CCC was moderate for the MSSE (CCC = 0.50), slight for the QOD–LTC (CCC = 0.18) and fair for the other instruments (CCC = 0.21–0.32).

Total scores between nurses and physicians in NHs differed significantly for the EOLD–SM, POS and QOD–LTC (Table 8). Physicians rated the quality as more favorable than did nurses for four of five instruments, with the difference being small to moderate ($d = 0.24$-0.58). Reliability was moderate for the EOLD–CAD (CCC = 0.59) and EOLD–SM (CCC = 0.48), and fair for the other instruments CCC = 0.25–0.40).

Discussion

We examined the validity and reliability of ten after-death instruments that are commonly used to assess family and professional caregivers' perspectives regarding QOC or QOD with dementia at the end-of-life. Of the instruments measuring QOC, the EOLD–SWC and FPCS showed better validity and internal consistency than the FPPFC, which in turn performed better than the FATE–S and the TIME. Of the QOD instruments, the EOLD–CAD and MSSE were most valid and internally consistent, followed by the EOLD–SM. The POS was not a valid measure to assess the QOD of dementia people in LTC settings. The QOD–LTC captures both QOC and QOD, and evidenced mean validity and internal consistency. The concordance between mean

Table 6 Reliability of instruments (and subscales) for instruments completed by family (N = 70[a]) and professional caregivers (N = 103)[a]

Measure	Internal consistency		CFA professionals[b]		
	Cronbach's α family	Cronbach's α Professionals	CFI	TLI	RMSEA (90% CI)
QOC					
FPCS		–	–	–	–
Resident care	0.85				
Family support	0.84				
Rooming	0.31				
Communication	0.88				
FPPFC	0.93				
TIME					
Physical comfort and emotional support	–				
Inform and promote shared decision making	0.58				
Encourage advance care planning	0.91				
Focus on individual	0.79				
Attend to emotional and spiritual needs of family	0.41				
Provide coordination and care	0.31				
Support self-efficacy of family	0.56				
Overall rating scale for patient focused, family centerd care	0.89				
QOD					
EOLD–CAD			0.98	0.98	0.08 (0.05–0.11)
Physical distress	0.64	0.64			
Dying symptoms	0.70	0.67			
Emotional distress	0.78	0.72			
Well-being	0.83	0.89			
QOC and QOD					
QOD–LTC			0.87	0.83	0.13 (0.10–0.16)
Personhood	0.70	0.75			
Preparatory tasks	0.43	0.66			
Closure	0.50	0.37			

Abbreviations of instruments are listed under Table 1

CFA Confirmatory factor analysis, *RMSEA* root mean square error of approximation, *CFI* comparative fit index, *TLI* Tucker–Lewis index, *SD* standard deviation

[a] Numbers were lower for instruments assessed by interview: the TIME (n = 63) and FPCS (n = 23 last decedents)

[b] No CFA was performed on family caregiver data because of the small number of cases

and individual scores given by family, physician and nurses was low for four of five instruments (the EOLD–CAD, EOLD–SM, POS, and QOD–LTC), likely reflecting the different perspectives, including different values and points of reference, of the family and professional caregivers. Thus, respondent type should be considered when selecting an instrument for use, as both perspectives offer valuable information unique to the situation [39]. However, to improve QOC and QOD, both perspectives are important. Therefore, for research and individual assessments, an integrated approach may be the best [40].

The correlations across instruments that measure the same construct were moderate to good. The correlation between the total scores of EOLD–CAD and MSSE was good ($r = -0.58$) and similar to the correlation Aminoff found for people with dementia ($r = -0.5$ to -0.8) [13]. Internal consistency was similar to literature reports for most instruments, except for the TIME for which it was slightly higher.

The correlation of the QOD instruments with the overall assessment was lower than for the QOC instruments. It is possible that ratings of QOD are more individualized than

Table 7 Instrument scores of family caregivers and professional caregivers ($N = 54$)

Measure (range)	Mean scores (SD)[a]		P value mean diff	d	CCC (95% CI)
	Family caregivers	Professional caregivers			
QOD					
EOLD–CAD (14–42)	30.8 (5.4)	32.9 (5.4)	0.01	0.39	0.32 (0.05; 0.55)
EOLD–SM (0–45)	27.6 (10.0)	34.4 (8.0)	<0.001	0.75	0.21 (−0.06; 0.46)
MSSE (0–8[b])	3.1 (1.8)	3.1 (1.8)	0.88	0.02	0.50 (0.25; 0.68)
POS (0–4)	1.1 (0.6)	1.0 (0.5)	0.82	0.18	0.23 (−0.09; 0.50)
QOC and QOD					
QOD–LTC (1–5)	3.2 (0.7)	3.6 (0.7)	<0.001	0.57	0.18 (−0.06; 0.41)

Abbreviations of instruments are listed under Table 1

CCC Concordance correlation coefficient, *CI* confidence interval; *d* effect size, *SD* standard deviation

[a] For the POS and MSSE, a lower score reflects a higher quality and for the other instruments a higher score signifies a higher quality

[b] The opinions about suffering out of different perspectives were not included in this comparison

Table 8 Instrument scores of Physicians and Nurses in NHs ($N = 33$)

Measure (range)	Mean scores (SD)[a]		P value mean diff	d	CCC (95% CI)
	Physicians	Nurses			
QOD					
EOLD–CAD (14–42)	31.8 (5.5)	30.5 (5.2)	0.14	0.24	0.59 (0.32; 0.77)
EOLD–SM (0–45)	32.2 (9.1)	26.5 (10.4)	<0.001	0.58	0.48 (0.12; 0.72)
MSSE (0–8[b])	3.4 (1.7)	4.1 (2.1)	0.08	0.37	0.40 (0.08; 0.64)
POS (0–4)	1.0 (0.4)	1.2 (0.6)	0.03	0.39	0.25 (−0.07; 0.54)
QOC and QOD					
QOD–LTC (1–5)	3.0 (0.6)	3.3 (0.6)	0.03	0.50	0.28 (−0.03; 0.56)

Abbreviations of instruments are listed under Table 1

CCC Concordance correlation coefficient, *CI* confidence interval, *d* effect size, *SD* standard deviation

[a] For the POS and MSSE a lower score reflects a higher quality; for the other instruments a higher score signifies a higher quality

[b] The opinions about suffering were not included in this comparison

are ratings of QOC, and so less well captured in standardized measures. For example, the presence of particular symptoms may be expected by some as part of the dying process, while others may find them very disturbing [41]. The QOD instruments assess frequency of symptoms whereas the overall assessment may include the degree to which symptoms are perceived as burdensome, or other constructs that are not reflected in the individual measures.

The mean scores of the prospective observations with the same instruments were more favorable than of the retrospective assessments. This difference may reflect the shorter time frame of prospective observations (average of five minutes versus last week of life). Further, the difference was largest when the retrospective assessment was performed by a family (EOLD–CAD and MSSE) or nurse (EOLD–CAD), and both generally assigned less favorable scores than physicians. The CCC for concordance between prospective observer and retrospective physician assessment of the EOLD–CAD was low but the total EOLD–

CAD scores were not significantly different, as found in previous work [39]. The prospectively assessed MSSE may compare better to the retrospectively assessed MSSE than the EOLD–CAD, because the MSSE may be more stable over time (i.e., it includes more stable items in addition to symptoms). Further, comparing slightly different constructs such as observed discomfort and quality of life in retrospect, a CCC of about 0.5 is expected [42]. Nevertheless, our results confirm the validity of the tested instruments in a sense that the retrospective assessments are at least to some degree based on prospectively observed experiences of dying.

Although sample sizes for some of our analyses were small, and the numbers were smaller for the FATE–S and FPCS which we added late, this study offers a well-rounded, in-depth examination of existent measures to assess the QOC and QOD. Exploratory analysis showed no systematic differences between respondents who did and did not complete the FATE–S and FPCS. A strength of our

study is that we included prospective observations and professionals' views, where as most studies are limited to retrospective family evaluations. Of note, Parker et al. recently published a systematic review of palliative care outcome measures used to assess the quality of palliative care provided in residential aged care facilities [43]. They identified eight of the same instruments evaluated by this analysis and no additional instruments were identified that fit our inclusion criteria [7]. Finally, we presented only internal consistency and CFA of reflective (sub)scales, although the differences between reflective (or "clinimetric") and formative (or "psychometric") measures is somewhat subjective. We chose to present internal consistency and CFA also when the nature of the measure was in question. A recently performed factor analysis [20] showed that the POS consists of two subscales and some solitary items, and therefore it is not useful to calculate the internal consistency for the whole POS instrument.

The psychometric properties of the instruments presented in this paper are generally consistent with findings in a parallel matched-design study of US family respondents, although there are some differences which will be addressed in detail in future work [44]; still the results of the Dutch study likely apply to other Western countries when using the English versions or properly translated versions. Further work will also consider usefulness and feasibility issues to help decide which instruments are best to use for measuring QOC and QOD for those dying with dementia in LTC settings.

References

1. Mitchell, S. L., Teno, J. M., Miller, S. C., & Mor, V. (2005). A national study of the location of death for older persons with dementia. *Journal of the American Geriatrics Society, 53*, 299–305.
2. Houttekier, D., Cohen, J., Bilsen, J., Addington-Hall, J., Onwuteaka-Philipsen, B. D., & Deliens, L. (2010). Place of death of older persons with dementia. A study of five European countries. *Journal of the American Geriatrics Society, 58*, 751–756.
3. Birch, D., & Draper, J. (2008). A critical literature review exploring the challenges of delivering effective palliative care to older people with dementia. *Journal of Clinical Nursing, 17*, 1144–1163.
4. van der Steen, J. T. (2010). Dying with dementia: What we know after a decade with research. *Journal of Alzheimer's Disease, 22*, 37–55.
5. Ferri, C. P., Prince, M., Brayne, H., et al. (2005). Global prevalence of dementia: A Delphi consensus study. *Lancet, 366*, 2112–2117.
6. Health Council of the Netherlands. (2002). Dementie. The Hague: Health Council of the Netherlands.
7. van Soest-Poortvliet, M. C., van der Steen, J. T., Zimmerman, S., Cohen, L. W., Munn, J., Achterberg, W. P., et al. (2011). Measuring the quality of dying and quality of care when dying in long-term care settings: A qualitative content analysis of available instruments. *Journal of Pain and Symptom Management.* doi:10.1016/j.jpainsymman.2011.02.018.
8. Volicer, L., Hurley, A. C., & Blasi, Z. V. (2001). Scales for evaluation of end-of-life care in dementia. *Alzheimer Disease and Associated Disorders, 15*(4), 194–200.
9. Volicer, L., Hurley, A. C., & Blasi, Z. V. (2003). Characteristics of dementia end-of-life care across care settings. *The American Journal of Hospice and Palliative Care, 20*(3), 191–200.
10. Kiely, D. K., Volicer, L., Teno, J., Jones, R. N., Prigerson, H. G., & Mitchell, S. L. (2006). The validity and reliability of scales for the evaluation of end-of-life care in advanced dementia. *Alzheimer Disease and Associated Disorders, 20*(3), 176–181.
11. Casarett, D. J., Pickerd, A., Bailey, A., et al. (2008). A nationwide VA palliative care quality measure: The family assessment of treatment at the end of life. *Journal of Palliative Medicine, 11*(1), 68–75.
12. U.S. department of veteran affairs (2011, July). Family assessment of treatment at the end-of-life (FATE) survey development. http://www.hsrd.research.va.gov/research/abstracts.cfm?Project_ID=2141694967
13. Vohra, J. U., Brazil, K., Hanna, S., & Abelson, J. (2004). Family perceptions of end-of-life care in long-term care facilities. *Journal of Palliative Medicine, 20*(4), 297–302.
14. Biola, H., Sloane, P. D., Williams, C. S., Daaleman, T. P., Williams, S. W., & Zimmerman, S. (2007). Physician communication with family caregivers of long-term care residents at the end of life. *Journal of the American Geriatrics Society, 55*, 846–856.
15. Teno, J. (2000). Toolkit of instruments to measure end of life care. After death bereaved family interviewer: NH version. http://www.chchr.brown.edu/pcoc.linkstoinstrumhtm.htm.
16. Teno, J. M., Clarridge, B., Casey, V., Edgman-Levitan, S., & Fowler, J. (2001). Validation of toolkit after-death bereaved family member interview. *Journal of Pain and Symptom Management, 22*(3), 752–758.
17. Aminoff, B. Z., Purits, E., Noy, S., & Adunsky, A. (2004). Measuring the suffering of end-stage dementia: Reliability and validity of the mini-suffering state examination. *Archives of Gerontology and Geriatrics, 38*, 123–130.
18. Hearn, J., & Higginson, I. J. (1999). Development and validation of a core outcome measure for palliative care: Palliative care outcome scale. *Qual Health Care, 8*, 219–227.
19. Brandt, H. E., Deliens, L., van der Steen, J. T., Ooms, M. E., Ribbe, M. W., & van der Wal, G. (2005). The last days of life of patients with and without dementia assessed with the palliative care outcome scale. *Palliative Medicine, 19*, 334–342.
20. Siegert, R. J., Gao, W., Walkey, F. H., & Higginson, I. J. (2010). Psychological well-being and quality of care: A factor-analytic examination of the palliative care outcome scale. *Journal of Pain and Symptom Management, 40*(1), 67–74.
21. Munn, J. C., Zimmerman, S., Hanson, L. C., et al. (2007). Measuring the quality of dying in long-term care. *Journal of the American Geriatrics Society, 55*, 1371–1379.

22. Gijsberts, M. J., van der Steen, J. T., Muller, M. T., & Deliens, L. (2008). [End-of-life with dementia in Dutch antroposofic and traditional nursing homes] Tijdschr Gerontol Geriatr, 39(6), 256–264. Dutch.

23. Schols, R., Schipper, R.-J., Brabers, A., & Schols, J. (2003). [The Mini-Suffering State Exam (MSSE) has been studied in a Dutch nursing home]. Tijdschrift voor Verpleeghuisgeneeskunde, 27, 14–18. Dutch.

24. Mitchell, S. L, Kiely, D. K., Jones, R. N., et al. (2006). Advanced dementia research in the nursing home: The CASCADE study. Alzheimer Disease and Associated Disorders, 20, 166–175).

25. Volicer, L., Hurley, A. C., Lathi, D. C., & Kowall, N. W. (1994). Measurement of severity in advanced Alzheimer's disease. Journal of Gerontology, 49(5), M223–M226.

26. van der Steen, J. T., Volicer, L., Gerritsen, D. L., Kruse, R. L., Ribbe, M. W., & Mehr, D. R. (2006). Defining severe dementia with the Minimum Data Set. International Journal of Geriatric Psycology, 21, 1099–1106.

27. Conroy, S., van der Cammen, T., Schols, J., van Balen, R., Peteroff, P., & Luxton, T. (2009). Medical services for older people in nursing homes—comparing services in England and the Netherlands. Journal of Nutrition, Health and Aging, 13(6), 559–563.

28. Hurley, A. C., Volicer, B. J., Hanrahan, P. A., Houde, S., & Volicer, L. (1992). Assessment of discomfort in advanced alzheimer patients. Research in Nursing and Health, 15, 369–377.

29. Warden, V., Hurley, A. C., & Volicer, L. (2003). Development and psychometric evaluation of the pain assessment in advanced dementia (PAINAD) scale. Journal of American Medical Directors Association, 4, 9–15.

30. Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). New Jersey: Lawrence Erlbaum.

31. Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45, 255–268.

32. Crawford, S. B., Kosinski, A. S., Lin, H. M., Williamson, J. M., & Barnhart, H. X. (2007). Computer programs for the concordance correlation coefficient. Computer Methods and Programs in Biomedicine, 88, 62–74.

33. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159–174.

34. Fayers, P. M., Hand, D. J., Bjordal, K., & Groenvold, M. (1997). Causal indicators in quality of life research. Quality of Life Research, 6, 393–406.

35. Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. Psychological Methods, 5, 155–174.

36. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55.

37. Terwee, C. B., Bot, S. D. M., de Boer, M. R., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 60(12), 34–42.

38. Muthén, L. K., & Muthén, B. O. (1998–2010). Mplus user's guide (6th ed.). Los Angeles, CA: Muthén & Muthén.

39. Teno, J. M., Landrum, J. M., & Lynn, J. (1997). Defining and measuring outcomes in end-stage dementia. Alzheimer Disease and Associated Disorders, 11(Suppl 6), 25–29.

40. van der Steen, J. T., Gijsberts, M. J. H. E., Knol, D. L., Deliens, L., & Muller, M. (2009). Ratings of symptoms and comfort in dementia patients at the end of life: Comparison of nurses and families. Palliative Medicine, 23, 317–324.

41. Shega, J. W., Houghham, G. W., Stocking, C. B., Cox-Hayley, D., & Sachs, G. A. (2008). Patients dying with dementia experience at the end of life and impact of hospice care. Journal of Pain and Symptom Management, 35, 499–507.

42. Schalkwijk, D., Verlare, L. R., Muller, M. T., Knol, D. L., & van der Steen, J. T. (2009). Het meten van kwaliteit van leven bij ernstig demente verpleeghuisbewoners: psychometrische eigenschappen van de QUALID-schaal. Tijdschrift voor Gerontologie en Geriatrie, 40(5), 184–192.

43. Parker, D., & Hodgkinson, B. (2011). A comparison of palliative care outcome measures used to assess the quality of palliative care provided in long-term care facilities: A systematic review. Palliative Medicine, 25(1), 5–20.

44. Cohen, L. W., van Soest-Poortvliet, M. C., van der Steen, J. T., de Vet, H. C. W., Reed, D., & Zimmerman, S. (2009). Differences in the end-of-life experience in the United States and Netherlands. Gerontologist, 49(suppl 2), 58.