Pain xxx (2006) xxx–xxx

PAIN

www.elsevier.com/locate/pain

The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia

Sandra M.G. Zwakhalen [a,*], Jan P.H. Hamers [a], Martijn P.F. Berger [b]

[a] *Department of Health Care Studies, Section of Nursing Science, Maastricht University, The Netherlands*
[b] *Department of Methodology and Statistics, Maastricht University, The Netherlands*

Received 17 May 2006; received in revised form 19 June 2006; accepted 26 June 2006

Abstract

In view of the need for valid, reliable, and clinically useful scales to assess pain in elderly people with dementia, this study evaluated the psychometric properties of translated versions of the PAINAD, PACSLAC, and DOLOPLUS-2 scales. In an observational study design, two raters simultaneously assessed the nursing home residents ($n = 128$) for pain during influenza vaccination and care situations. The PACSLAC was valued as the most useful scale by nurses. Cronbach's alpha was high (>.80) for the total scale at T2 and T3 and adequate for the 'Facial expression' and 'Social/personality/mood' subscales. IC scores for the 'Activity/body movement' and 'Physiological indicators/eating/sleeping changes/vocal behaviors' subscales were low. It demonstrated good validity and reliability, although the scale should be further refined. This refinement should increase homogeneity. The PAINAD showed good psychometric qualities in terms of reliability, validity, and homogeneity (α ranged .69–.74 at T2 and T3) (except for the 'Breathing' item). The PAINAD scale had lower scores for clinical usefulness in this sample. The Dutch version of the DOLOPLUS-2 was considered more difficult to use but showed acceptable psychometric qualities in terms of the issues assessed, except for the 'psychosocial reactions' subscale. IC of the DOLOPLUS were adequate for the total scale (α ranged .74–.75) and almost all subscales (α ranged .58–.80). Findings of this study provide evidence of validity and reliability of the three pain assessment scales. Now that a pain scale is available, future studies also need to focus on its implementation in nursing practice.

Keywords: Nursing homes; Dementia; Elderly residents; Pain assessment; Psychometric

1. Introduction

Ageing is known to be associated with high prevalences (up to 80%) of persistent pain among those living in nursing homes (Ferrell, 1995; Weissman and Matson, 1999; Lansbury, 2000; Helme and Gibson, 2001). Assessing pain is crucial, as inadequate assessment could mean that pain remains under-detected and under-treated, with negative effects on quality of life. However, measuring pain is often extremely difficult, especially in people with dementia. Several factors might contribute to the complexity of this assessment. First, dementia is a syndrome characterized by progressive decline in cortical functions (Farrell et al., 1996). This may reduce a person's ability to interpret and report pain. Moreover, behavioral and psychological repertoires of elderly people with dementia can be very heterogeneous. Just as in cognitively impaired children, behavior that is typically associated with pain (like yelling) may well appear in non-painful situations, or behavioral limitations may mask expressions of pain (McGrath et al., 1998). In addition, increasing age is often associated with increasing somatic and physical problems, limiting people in their ability to express pain.

In the last decade, several observational scales have been developed to measure pain in (severely) demented

* Corresponding author. Tel.: +31433884083; fax: +31433884162.
E-mail address: S.Zwakhalen@zw.unimaas.nl (S.M.G. Zwakhalen).

0304-3959/$32.00
doi:10.1016/j.pain.2006.06.029

people. A review of the literature identified 12 observational pain scales, whose psychometric qualities have been evaluated (Zwakhalen et al., 2006). Findings of this review indicate that most observational scales are still under development and show moderate psychometric qualities. Most scales lack validity, reliability, and clinical usefulness. Nevertheless, the findings demonstrated that PAINAD, PACSLAC, and DOLOPLUS-2 showed promising psychometric qualities, though they required further testing.

Until now, no Dutch pain assessment scale is available to reliably and validly measure pain in nursing home settings. The number of newly developed behavioral scales has grown rapidly and in view of this growth, it seemed useful to adapt and improve these scales for use in The Netherlands, instead of developing a new tool. Therefore, three scales were translated. A backward–forward method was used in the translation procedure to adapt these scales for use in Dutch nursing home settings. Translated scales were finally adjusted based on the results of pre-testing to investigate nurses' opinion on their clinical usefulness (Zwakhalen et al., in preparation). The final versions of the translated scales were used in the present study.

This study was thus initiated to evaluate the psychometric properties of three translated pain assessment scales (PAINAD, PACSLAC, and DOLOPLUS-2) in nursing home practice, using potentially painful situations.

More specifically, this study addressed the following research questions:

- What items of the translated PAINAD, PACSLAC, and DOLOPLUS-2 scales are most frequently used to assess pain in elderly nursing home patients with dementia?
- Are the translated versions of the scales and subscales internally consistent?
- Are the translated versions of the scales reliable (in terms of inter-rater and intra-rater reliability)?
- Do the translated versions of the scales show good construct validity using a known-groups technique?
- Do the translated versions of the scales show good congruent validity?
- How do nurses rate the clinical usefulness of the scales?

2. Methods

2.1. Design

An observational study design was used to answer the research questions. Observations were carried out with patients at rest and during potentially painful situations/interventions. All patients were observed while receiving their influenza vaccination and a selection of patients were also observed at a patient-specific pain moment (like care, washing or mobilization). In prior pain research (e.g., Oberlander et al., 1999; Breau et al., 2001), injections have been successfully used as potentially painful stimuli.

2.2. Participants

The study involved 144 nursing home patients, including 128 demented patients of 12 psycho-geriatric (PG) wards and 16 somatically ill nursing home patients of one ward receiving their annual planned influenza vaccination. This group of somatically ill patients was included to serve as a control group for pain intensity scores (using a VAS only) during the influenza injection.

To be included, elderly residents with dementia had to have been living in the nursing home for at least 4 weeks prior to the influenza vaccination. Patients were at least 60 years old and had not undergone major environmental changes in the last month. Excluded were patients with acute medical illness or a purely psychiatric disorder and patients with Korsakov's syndrome. The type of dementia syndrome as well as the severity of the impairment varied.

Twelve nurses of three nursing homes in The Netherlands participated in this study.

2.3. Measures

A *data sheet* was used to collect demographic and additional information about medication use (analgesics and psychotropics), environmental changes during the last month, and type of dementia diagnoses. Information was obtained from patients' medical and nursing records.

The patients' cognitive status was evaluated using the *Minimum Data Set (MDS) Cognitive Performance Scale (CPS)* (Morris et al., 1994). The CPS combines five MDS items relating to cognitive functioning, eventually yielding seven categories of cognitive impairment (Hartmaier et al., 1995). The MDS-CPS has been validated against the Mini-Mental State Examination (MMSE) (Folstein et al., 1975) and showed substantial agreement (Hartmaier et al., 1995).

The present study used Dutch versions of three selected pain assessment scales (*DOLOPLUS-2, PAINAD, and PACSLAC*).

The *DOLOPLUS-2* scale by Wary (1999, 2001) consists of 10 items covering the somatic, psychomotor, and psychosocial impacts of pain. Each of the 10 items can be described at one of four different levels – rated from 0 to 3 – representing increasing intensity of pain (Lefebvre-Chapiro and Doloplus group, 2001). A score of at least 5 out of 30 (the maximum pain score) is considered to indicate pain. Several studies have been conducted in geriatric centers and palliative care units to validate the scale, investigating test–retest reliability, concurrent validity, and inter-rater reliability (Michel et al., 2000; Serbouti et al., 2002). According to the authors, the findings showed convergent validity of the DOLOPLUS-2 with a VAS and that DOLOPLUS-2 demonstrated good sensitivity. There was satisfactory stability on the retest. An inter-rater correlation test between two physicians showed no significant difference ($p < 0.001$) and high levels of internal consistency ($\alpha = 0.82$) were found. Howev-

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

er, the publications about the DOLOPLUS-2 provide almost no concrete information about correlation coefficients (inter-rater reliability, test–retest reliability) and about the determination of cut-off-scores and the impairment level of the participants (Zwakhalen et al., 2006).

The *Pain Assessment Checklist for Seniors with Limited Ability to Communicate (PACSLAC)* by Fuchs et al. (Fuchs et al., 2002; Fuchs-Lacelle and Hadjistavropoulos, 2004) consists of 60 items covering four subscales. PACSLAC, which is still under construction, has good content validity. The PACSLAC developers collected items specifically indicating pain in elderly people with dementia (Zwakhalen et al., 2006). PACSLAC has been the subject of a partial, preliminary validation study. High levels of internal consistency were found for the total scale ($\alpha = 0.82$–0.92), although Cronbach's α values for the subscales were lower (.55–.73). Therefore, the developers of the PACSLAC scale recommended use of the total score only (Fuchs-Lacelle and Hadjistavropoulos, 2004). The PACSLAC total score seemed to discriminate between painful, calm, and distressing events. Correlations calculated between global intensity ratings and PACSLAC scores were moderate ($r = .39$–$.54$). Correlations among the subscales suggest that although the checklist measures a unified construct, the subscales are sufficiently discriminatory (Fuchs-Lacelle and Hadjistavropoulos, 2004).

The *Pain Assessment IN Advanced Dementia Scale (PAINAD)* by Warden et al. (Warden et al., 2003) can be described as a modification of the Discomfort Scale for Patients with Dementia of the Alzheimer Type (DS-DAT) by Hurley et al. (1992). The current version consists of five items with three response options scored from 0 to 2. The instrument has been tested in a residential setting (dementia care ward) involving 19 severely impaired patients. Although Warden et al. reported high levels of inter-rater reliability (Pearson $r = .82$–$.97$), internal consistency was below $\alpha < .70$. The scale showed evidence of construct validity. The tool correlated well with the DS-DAT ($r = .76$), a VAS for discomfort ($r = .81$), and a VAS for pain ($r = .75$). Pain scores were found to be lower during pleasant than during aversive activities and scores differed before and after pain modification (Lane et al., 2003; Warden et al., 2003).

In our study, a *Visual Analogue Scale (VAS)* was used by nurses and an independent rater (rater 1) to estimate the pain intensity. A VAS is a 100 mm line on which pain can be marked with anchors at either end. The left-hand anchor was labeled 'no pain at all' and the right-hand anchor was labeled 'pain as bad as it could possibly be' (Jensen and Karoly, 2001). VAS is frequently used in pain studies as a self-report scale or for proxy reports of pain.

We used two self-report scales to assess pain in our elderly patients with dementia, namely a *Verbal Rating Scale (VRS)* (Jensen and Karoly, 2001) and the *Color Analogue Scale (CAS)* developed by McGrath et al. (McGrath et al. (1996)). The CAS is a colored variant of the VAS developed as a pain intensity scale for children. The VRS consists of a list of adjectives describing various levels of pain intensity (Jensen and Karoly, 2001) using a 4-point scale ranging from 0 (no pain) to 3 (severe pain), and was also used as a self-report scale. CAS and VRS have demonstrated their usefulness and showed adequate to high completion rates in previous research on pain assessment in elderly people with various types of dementia (Scherder et al., 2003; Closs et al., 2004).

Information on pain was also gathered using the Residents Assessment Instrument (RAI) Minimum Data Set (MDS), of which we applied the items on pain frequency (J2a), pain intensity (J2b), and location of pain (J3) (Morris et al., 1994).

To assess the usefulness of the three scales (DOLOPLUS, PACSLAC, and PAINAD), nurses were asked which scale they considered most useful and which scale they preferred. A 10-point scale was used for scoring the clinical usefulness. Finally, nurses had the opportunity to comment on the scales (on aspects like confusing items, overlapping information, and scoring difficulties).

2.4. Procedure

Two raters (rater 1 and rater 2) simultaneously assessed the patients' pain at rest (T1) and during vaccination (T2). A selected group of patients was also assessed during a specific moment (T3). Observations of the first two moments lasted for approximately 2 min and were conducted on the same day. To avoid differences in scores due to developments over time, the time interval between T2 and T3 was minimized, and was mostly shorter than 3 weeks.

Demographic information was gathered by the nurse when residents were at rest (T1). At T1, T2, and T3, the VAS and pain assessment scales (PACSLAC and PAINAD) were scored in random order immediately after the patient had been observed. Because DOLOPLUS-2 does not represent pain experience at a specific moment but reflects the progression of pain experienced, the DOLOPLUS-2 scale was scored last, at a separate time after the observations. Because the DOLOPLUS-2 cannot be used without in-depth knowledge of the patient, the DOLOPLUS-2 was the only scale that was not scored simultaneously (in the absence of rater 1 scores). Patients were questioned about their current pain intensity immediately after all moments, using CAS and VRS at random. At T3, participants underwent a variety of specific potentially painful interventions (like care activities, washing or mobilization). Patients who, in the opinion of the nurses involved, were not expected to be in pain at specific moments were not assessed a third time.

Clinical usefulness was assessed last, after the nurses had used all scales for all participants at their ward. An overview of the measures used at particular moments by the raters is presented in Table 1.

2.4.1. Videotaping

Video recordings were obtained by a research assistant for every patient, after approval had been obtained from legal guardians. The observational moments were videotaped, as part of the study, to test the inter-rater reliability. Video recordings were made from an estimated distance of 2.5 m and were conducted in the patients' customary environment. Videotaping was once interrupted in a situation where the patients showed resistance to taping.

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

Table 1
Overview of measures used by the raters at various moments

Measures	Live rating at the bedside						Retrospective rating
	Moment at rest (T1)		Moment of injection (T2)		Specific moment (T3)[a]		Video recording (T4)
	Nurse	Rater 1	Nurse	Rater 1	Nurse	Rater 1	Rater 1
Data sheet incl. MDS-CPS	x						
VAS pain	x	x	x	x	x	x	x
PAINAD	x	x	x	x	x	x	x
DOLOPLUS	x				x		
PACSLAC	x	x	x	x	x	x	x
Clinical usefulness evaluation[b]					x		

[a] This observational moment was not included for all patients; selection was based on nurses' estimation of pain potentially occurring at a specific moment.
[b] Usefulness was assessed after all scores had been obtained (at T2 or T3).

2.5. Ethical considerations

The study was approved by the Medical Ethics Committee of the University Hospital Maastricht and the University of Maastricht. Permission to conduct the study was also obtained from the managing directors of the nursing homes. Approached nurses participated on a voluntary basis.

Before participation, registered legal guardians of the residents provided written informed consent. One of the response options provided on the informed consent form was participating in the study but refusing videotaping (with 26.6% of the legal guardians refusing videotaping).

2.6. Analyses

Descriptive statistics were generated for the characteristics of the respondents, as well as for examining the clinical usefulness. Frequency of endorsement ratings was also computed using descriptive statistics to examine the frequency with which various items were used.

Internal consistency (IC) analyses were carried out, the IC of the scales and subscales being determined using Cronbach's α. Inter- and intra-rater reliability for the pain scores was measured by intra-class correlation coefficients (ICCs). The two-way mixed consistency method was used to examine the intra-rater reliability, since these scores focus on results within the study. In order to allow generalization of results, the two-way random absolute agreement method was used to examine the inter-rater reliability, which compensates for an extra source of variance due to differences between raters.

Construct validity was examined using the known-groups technique (Polit and Beck, 2004). On the assumption that higher pain scores would lead to higher sum scores on the PAINAD and PACSLAC, mean scores were calculated. Pain versus non-pain groups were created (assuming that pain at T2 is higher than at T1 and based on VAS scores, viz. VAS = 0 versus VAS $\geqslant$ 30). To examine the construct validity of the DOLOPLUS-2, pain versus non-pain group were created in slightly different manner because DOLOPLUS-2 scores cannot be collected specifically in a context like an observational moment during care. Therefore groups were created based on information about 'no pain' versus 'daily pain' derived from the MDS scale (pain frequency).

Finally, in order to determine congruent validity, Pearson correlations were computed between the measures. Congruent validity was implied by high correlations between PAINAD and PACSLAC, since the instruments both claim to be adequate pain measures. A moderate to high correlation was expected with the VRS and VAS scores produced by the various raters.

3. Results

3.1. Participants

A total of 128 nursing home patients with dementia participated in this study, 21.9% ($n = 28$) men and 78.1% ($n = 100$) women. Ages ranged from 60 to 96 years, with a mean age of 82.4 (SD = 6.8). Sixteen somatically ill patients with an average age of 78.1 years also participated in this study, 5 men and 11 women.

Nursing staff ($n = 12$), mostly female ($n = 10$), ranged in age from 18 to 51 years (mean = 34.3: SD = 9.6). The mean number of years of experience in nursing was 10.2 (SD = 5.2). Table 2 presents further information on the participants' characteristics.

Analyses focused first on detailed information on the pain situation. Based on descriptive information derived from the MDS items on pain, 25% of the psycho-geriatric (PG) residents experienced pain on a daily basis, while 17.2% of the residents experienced pain less frequently than daily. Perceived pain intensity scores showed that of those residents frequently experiencing pain (42.2%), 14.1% had only mild pain, 22.7% experienced moderate pain, and 5.5% experienced unbearable pain now and then. MDS findings indicated that the most frequently mentioned pain locations were joints (23.4%), other places (10.9%), soft tissues (7.8%), back (3.9%), bones (3.1%), and hips (3.1%). Further investigation of the MDS item on pain frequency showed that almost 35% of the people who suffered daily pain received no pain medication at all.

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

Table 2
Descriptive information on the nursing home residents and nurses included in the study

Characteristics of somatic residents (n = 16)		
Male/Female		5/11
Age in years (SD)		78.1 (10.6)
Characteristics of PG residents (n = 128)		
Male/Female		28/100
Age in years (SD)		82.4 (6.8)
Severity of the impairment n = (%)		
Mild, CPS< = 2		28 (21.9%)
Moderate to moderately severe, CPS = 3 or 4		36 (28.1%)
Severe to very severe, CPS = 5 or 6		61 (47.7%)
Unknown, missing		3 (2.3%)
Dementia diagnosis n = (%)		
Alzheimer's disease		41 (32.0%)
Vascular dementia		24 (18.8%)
Other (e.g., Parkinson's disease, frontal lobe)		7 (5.5%)
Mixed (Alzheimer's/vascular)		5 (3.9%)
Unknown		51 (39.8%)
Medication (analgesic and psychotropic) n = (%)		
Analgesic	None	85 (66.4%)
	Regular	32 (25.0%)
	When necessary	8 (6.3%)
	Daily and when necessary	2 (1.6%)
Psychotropic	None	58 (45.3%)
	Regular	65 (50.8%)
	When necessary	1 (0.8%)
	Daily and when necessary	3 (2.3%)
Characteristics of nurses (n = 12)		
Male/Female		2/10
Age in years (SD)		34.3 (9.6)
Years of experience in nursing (SD)		10.2 (5.2)
Educational level		
Nurses (RN)/caregivers		5/7

3.2. Observed moments

A total of 128 participants with a dementia syndrome were observed on 290 occasions, including moments of rest (T1, $n = 128$), immediately after the influenza vaccination (T2, $n = 127$), and immediately after a patient-specific moment of potential pain (T3, $n = 35$). Most T3 observational moments were while patients were being washed in the morning ($n = 19$). Other T3 moments included transfer situations during care ($n = 3$), manipulation of the hand ($n = 4$), and wound care ($n = 1$).

Because most observations and ratings were carried out simultaneously by the nurses and rater 1, a total of 573 (seven missing) scores were gathered. The data were not normally distributed, since scores were frequently clustered around zero, especially at T1 and T2.

3.3. Frequencies of endorsed items and internal consistency

Item usage was assessed by calculating frequency of endorsement rates for items of the pain assessment scales for participants with a VAS greater than or equal to 30 ($n = 53$ moments). Given the fact that rater 1 was the consistent factor in all patient ratings, these VAS ratings were used.

The most frequently used items of the PACSLAC and PAINAD scales for patients with a VAS scored by rater 1 as $\geqslant 30$ are listed in Table 3.

In a painful situation (VAS $\geqslant$ 30), 28 of the 60 items of the translated version of the PACSLAC were not used for over 90% of the study participants, either by rater 1 or by the nurses who scored at the bedside. This indicates that these items could be 'less likely' candidates for inclusion in a general pain instrument for the target group of elderly people with dementia. For over 80% of the study participants, both raters did not use the item 'breathing' in PAINAD in painful situations (VAS $\geqslant$ 30).

For each of the scales (PACSLAC, PAINAD, and DOLOPLUS-2) and the subscales of the PACSLAC, Cronbach's α was calculated to examine the homogeneity of the scale items. The resulting values are shown in Table 4.

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

Table 3
Most frequently used items of PACSLAC and PAINAD (VAS rater 1> = 30)

Scale	Items	M2 % scored by rater 1	M3 % scored by rater 1
PACSLAC	Tighter face	96.4	100
	Change in eyes	96.4[a]	78.3
	Pain expression	82.1	69.6
	Frowning	75	78.3
	Creasing forehead	75[a]	65.2
	Specific sound or vocalisation for pain	71.4	65.2
	Grimacing	46.4	69.6
	Moaning and groaning	46.4	69.6
	Pulling away	46.4	47.8
	Opening mouth	42.9[a]	43.5
	Mumbling	32.1[a]	17.4
	Not wanting to be touched	28.6	43.5
	Upset	28.6	21.7
	Restless	25	30.4
	Flushed /red face	25	21.7
	Flinching	25	4.3
	Anxious	21.4	39.1
	Clenching teeth	17.9	43.5
	Stiff/rigid	14.3	30.4
PAINAD	Facial expressions	92.9	95.7
	Body language	89.2	95.7
	Negative vocalization	75	92.3

[a] Also frequently present (>15%) during a non-painful situation (VAS = 0).

The lowest internal consistency (IC) scores were found for the 'activity and body movement' subscale (ranging from .40 to .57) and the 'physiological indicators/eating/sleeping changes/vocal behaviors' subscale (ranging from .20 to .43), both of the PACSLAC. Although Cronbach's α for the total scale was high, it is positively influenced by the large number of scale items ($n = 60$ items).

Overall analyses showed adequate levels of IC for the PAINAD instrument at T2 and T3, ranging from .69 to .74. Corrected item–total correlations showed that the 'breathing' item scored persistently low at all moments (ranging from −.51 to .12).

IC scores of the DOLOPLUS-2 were adequate for almost all subscales (ranging from .58 to .80) and for the total scale (ranging from .74 to .75). Cronbach's α (.58–.63) was moderate for the 'psychosocial reactions' subscale but corrected item–total correlations never scored below .20.

3.4. Reliability of the translated scales

To examine inter- and intra-rater reliability, ICCs were calculated for the VAS, PAINAD, and PACSLAC. Inter-rater reliability was calculated between rater 1 and the nurses (mean scores) and was found to be high for all measures, including the VAS. Compared to the two behavioral scales, the VAS scored slightly lower. Furthermore, irrespective of the measure used, agreement was less perfect at T1 (moment of rest). Further analyses examined the differences between pairs of raters/observers. These analyses showed some remarkable variability in the inter-rater reliability between individual pairs of raters. Examining T2 specifically, inter-rater agreement between scoring pairs on the VAS ranged from −.32 to .97, those on the PAINAD from .00 to .95, and those on the PACSLAC total scale from .39 to .97.

Intra-rater reliability was assessed by comparing the scores allocated by rater 1 at the bedside with those allo-

Table 4
Internal consistency (Cronbach's α) of the pain assessment scales

	Rater 1			Nurses			Somatic patient
	T1	T2	T3	T1	T2	T3	T2
PAINAD total ($n = 5$ items)	0.48 (127)	0.69 (126)	0.72 (34)	0.59 (127)	0.74 (126)	0.72 (32)	
PACSLAC total ($n = 60$ items)	0.62 (127)	0.82 (126)	0.81 (34)	0.64 (127)	0.80 (126)	0.84 (33)	
Facial expressions ($n = 13$ items)	0.44	0.73	0.7	0.46	0.57	0.69	
Activity/body movement ($n = 20$ items)	0.1	0.4	0.55	0.28	0.49	0.57	
Social/personality ($n = 12$ items)	0.56	0.68	0.73	0.65	0.65	0.76	
Physiological/eating/sleeping changes/Vocal behaviors ($n = 15$)	0.32	0.43	0.28	0.12	0.37	0.2	
DOLOPLUS-2 ($n = 10$ items)	x	x	x	0.75 (89)	x	0.74 (26)	
Somatic reactions ($n = 5$)	x	x	x	0.7	x	0.63	
Psychomotor reactions ($n = 2$)	x	x	x	0.8	x	0.77	
Psychosocial reactions ($n = 3$)	x	x	x	0.63	x	0.58	
Mean VAS score somatic NH patient (min. 0–max. 100)		12.5 (0–36)					10.7 (0–31)
Mean VAS score PG NH patient (min. 0–max. 100)	0.9 (0–23)	17.1 (0–85)	36.3 (0–83)	1.0 (0–19)	13.9 (0–85)	35.2 (0–84)	

Abbreviations: PG, psycho-geriatric; NH, nursing home.

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

cated by re-scoring the same moments based on video recordings ($n = 29$ scenes). Intra-rater reliability scored high for all measures included. Reliability scores are presented in Table 5.

3.5. Validity of the translated scales

To examine the ability of the translated scales to detect pain, total scores for different pain moments were compared, based on the assumption that more pain would increase the number of observed behaviors and hence the total scores on the PAINAD and PACSLAC scales. Pain versus non-pain groups were created. Higher VAS scores at T2 and T3 confirmed that pain was more prevalent at these moments than at T1 (at rest). In addition, pain versus non-pain groups were also created on the basis of the patients' self-reports, rated by VRS.

Approximately half of the participants were able to use the VRS. The ability of the nursing home residents to complete a VRS differed between T1, T2, and T3. Completion rates were lowest at T1, when 78.6% ($n = 21$) of the people with a mild cognitive impairment were able to complete the VRS, compared to 61.1% ($n = 22$) of the moderately impaired and 26.2% ($n = 16$) of the severely impaired residents. At T2, 82.1% ($n = 23$) of mildly impaired, 77.1% ($n = 27$) of moderately impaired, and 37.7% ($n = 23$) of severely impaired residents were able to complete the VRS. At T3, 70% ($n = 7$) of mildly impaired, 55.6% ($n = 5$) of moderately impaired, and 12.5% ($n = 2$) of severely impaired residents were able to complete the VRS to report their pain intensity.

Results presented in Table 6 show a consistent upward trend, confirming the above hypothesis and supporting construct validity.

Since DOLOPLUS-2 scores cannot be easily gathered after a specific observational moment, we examined the construct validity of the DOLOPLUS-2 scale by comparing total DOLOPLUS-2 scores between a 'non-pain group' and a 'pain group'. Groups were created based on information about 'no pain' versus 'daily pain' derived from the MDS scale (pain frequency), with mean total scores in the 'daily pain' group obviously higher (mean 9.8; SD 6.0; range 2–23) than those for the 'no pain group' (mean 5.1; SD 3.9; range 0–16).

Table 6
Mean total score on PAINAD and PACSLAC (rater 1)

	PACSLAC	PAINAD
	Mean (min; max)	Mean (min; max)
VAS < 30	3.2 (0; 15)	0.6 (0; 4)
VAS 30–60	11.2 (6; 22)	3.8 (1; 6)
VAS > 60	15.0 (10; 24)	5.9 (3; 8)
M1 at rest	2.6 (0; 11)	0.4 (0; 4)
M2 injection	5.5 (0; 22)	1.5 (0; 7)
M3 specific moment	9.6 (1; 24)	3.6 (0; 8)
VRS no pain	2.1 (0; 10)	0.3 (0; 3)
VRS minor pain	5.6 (1; 14)	1.6 (0; 5)
VRS moderate	10.6 (6;13)	3.3 (1; 6)
VRS severe	13.4 (6; 20)	4.7 (2; 7)

Congruent validity was assessed by comparing the scores of the three translated behavioral pain scales with those from other pain measures (VAS scored by rater 1, VAS scored by nurse, and VRS). Pearson correlations between PACSLAC, PAINAD, and other pain measures were all significantly positive. The magnitude of the correlations ranged from $r = .29$ to $r = .89$. The correlations were highest between the PAINAD scores and the VAS scores by rater 1 at the bedside. The correlation between PACSLAC and PAINAD scores was .85. As expected, low scores were found for the DOLOPLUS-2 compared to other pain measures. Findings confirm the difference between DOLOPLUS-2 as a pain measure and PAINAD and PACSLAC. Table 7 presents the congruent validity results.

3.6. Clinical usefulness

Finally, the nurses' ratings of clinical usefulness (scored on a 10-point scale) showed that 75% preferred

Table 5
Inter- and intra-rater reliability of the PAINAD, PACSLAC, and VAS scores

	Inter-rater[a]			Intra-rater[b]
	T1	T2	T3	
PAINAD total ($N = 5$ items)	0.75	0.85	0.81	0.89
PACSLAC total ($N = 60$ items)	0.93	0.95	0.96	0.86
Facial expressions ($N = 13$ items)	0.95	0.96	0.91	0.92
Activity/body movement ($N = 20$ items)	0.77	0.85	0.92	0.72
Social/personality ($N = 12$ items)	0.91	0.94	0.95	0.89
Physiological/eating/sleeping changes/vocal behaviors ($N = 15$)	0.90	0.91	0.83	0.89
VAS score	0.69	0.86	0.78	0.85

[a] Two-way random absolute agreement intra-class correlation.
[b] Two-way mixed consistency intra-class correlation.

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

Table 7
Pearson correlation between scales

	VAS rater 1	VRS	VAS nurse	VAS video	PACSLAC	PAINAD	DOLOPLUS
VAS rater 1	1	0.86	0.87	0.87	0.8	0.89	0.29
VRS		1	0.85	0.69	0.81	0.81	0.36
VAS nurse			1	0.76	0.72	0.81	0.33
VAS video				1	0.86	0.79	–
PACSLAC					1	0.85	0.29
PAINAD						1	0.34
DOLOPLUS							1

Correlations are significant at the 0.01 level (two-tailed).

the PACSLAC to measure pain in elderly residents with dementia. Mean scores were highest for the PACSLAC (mean 7.0; SD 0.5), while the PAINAD scored 5.89 (SD 1.7). The lowest usefulness scores were reported for DOLOPLUS-2 (mean 5.6; SD 2.2).

PAINAD and PACSLAC were considered user-friendly and not time-consuming by the participating nurses. Once they were used to the scale, they could assess patients within a few minutes.

Qualitative information was gathered from the participating nurses who used the scales in this study. The following comments nurses wrote on the clinical usefulness evaluation form reveal some additional important information:

- "PACSLAC is comprehensive, significant, and specifies behavior. Items guide you towards certain possible pain cues. The current translated version of the PACSLAC contains many items. Several items are superfluous and other items overlap. PACSLAC is feasible for all nurses, no matter what their educational level is".
- "Although PAINAD is brief and well-structured, in order to detect pain, I expect a scale with specific cues. PAINAD is too concise".
- "DOLOPLUS-2 provides a more general view. A clear manual is provided. The scale is difficult to score and interpret. Its is questionable whether all items of the DOLOPLUS are relevant to detect pain. The psychosocial items in particular are difficult to interpret as solid specific pain behavior. Other causes, like the dementia itself, could explain a change in psychosocial behavior. This is also a possible problem with a few PACSLAC items".

4. Discussion

The findings of this study provide evidence on the validity and reliability of three pain scales (PAINAD, PACSLAC, and DOLOPLUS-2) for nursing home residents with dementia.

Our main conclusions from the findings are as follows:

- The Dutch version of PACSLAC was valued as the most useful scale by care providers. It demonstrated good validity and reliability, although the scale should be further refined. This refinement should increase the homogeneity of the scale and subscales.
- The translated version of the PAINAD instrument showed good psychometric qualities in terms of homogeneity (except for the 'Breathing' item), reliability, and validity. The PAINAD scale had lower scores for clinical usefulness in this sample.
- The Dutch version of the DOLOPLUS-2 was considered more difficult to use but showed acceptable psychometric qualities in terms of the issues assessed, except for the 'psychosocial reactions' subscale.

4.1. Reliability

Cronbach's α for newly developed scales should be above .70 (Nunnally, 1978). Our findings showed adequate levels of internal consistency for the PAINAD scale. Cronbach's α ranged from .69 to .74 at T2 and T3, moments when participants were more likely to be in pain. 'Breathing' was the poorest scoring item of the PAINAD, with a corrected item–total correlation below .2. Using the usual rule of thumb which says that the corrected item–total correlation should have a value of at least .20 (Streiner and Norman, 2003), this item could be discarded to increase Cronbach's α by approximately .50. Warden et al. (2003) found similar results but decided not to discard this item because of the threat of respiratory diseases and the fact that changes in respiration have been mentioned in reaction to acute pain.

Although Cronbach's α was good (>.80) for the total PACSLAC scale at T2 and T3 and adequate for two subscales, IC scores of the 'activity/body movement' and 'physiological indicators/etc.' subscales were low. The fact that many of the PACSLAC items were frequently not used and the item–total correlation was below .20 indicates that the number of items could be reduced. If items of the translated scale were discarded based on the following criteria: (I) "(not) used in over 90%" and (II) item–total correlation below .20, then a number of 29 items would remain. Item reduction would

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

clearly increase the internal consistency. If this item reduction is indeed implemented, we suggest performing an additional factor analysis. A few items, like 'refusing medication', were probably not observed due to the data collection method used.

Internal consistencies of the DOLOPLUS were adequate for the total scale (ranged .74–.75) and almost all subscales (ranged .58–.80). Lowest internal consistencies were found for the 'psychosocial reactions' subscale. Hølen et al. (2005) recently translated the DOLOPLUS-2 into Norwegian and pilot-tested it in patients with cognitive impairment. Their analyses showed that the psychosocial category in their study was also the weakest component.

The observational methodology used in our study was more adequate for evaluating the psychometric qualities of the PAINAD and the PACSLAC than for the DOLOPLUS. As a result, we were unable to determine the intra- and inter-rater reliability of the DOLOPLUS. Our findings showed good inter- and intra-rater agreement for the PAINAD and PACSLAC. Closer examination revealed that many pain scores still clustered around 0, which may have influenced the agreement. Inter-rater agreement was mostly lower at T1 than at T2 and T3. This confirms that it is evidently harder to estimate a person's pain at rest or in a clinically normal situation without a specific stimulus.

4.2. Validity

The scales were able to detect differences between patients who were in pain or not in pain. Discriminating between pain and non-pain events is one of the most essential aspects of a useful tool. Our analysis of congruent validity found moderate to high correlations (.72–.89) between the VAS scores by the various raters and the behavioral scales. These high correlation rates between VAS and PAINAD are consistent with the result of a study by Warden et al. (2003). In a validation study on pain measures (recalling pain), Leong et al. (2006) found high correlations (Kendall's $\tau = .842$) between PAINAD and nurses' pain reports. By contrast, other studies found lower correlations between self-report and behavioral scales. Hadjistavropoulos et al. (2000) found that their self-report measure was not related to the non-verbal scales and expressed the opinion that both measures assessed different parameters of pain experience. The high correlations found in our study could be due to the fact that the scales were tested in a controlled situation with an acute stimulus. Nurses were not blinded to the intervention (influenza injection). Another aspect may also be that we used the Pearson correlation coefficient, a liberal measure (Streiner and Norman, 2003) which might partially explain the high correlations.

A significant number of patients were able to report pain by means of a VRS. Higher VRS scores were consistent with higher total scores on the observational scales PAINAD and PACSLAC. Given the high correlations between patients' VRS and proxies' VAS scores, the results of our study suggest that proxies adequately estimated the pain intensity of persons in a controlled situation. This does not, however, resolve some of the doubts often voiced about self-reports, which mostly relate to construct validity and reliability. Jensen concluded that the major validity question concerning self-report pain scales involves the extent to which self-report measures actually reflect pain experience (Jensen, 1997). Craig et al. (2001) noted that the tendency to trust non-verbal information more than self-report is widespread; "Non-verbal expressions will be more spontaneous and less subject to purposeful manipulation than self-report". Scherder et al. (2005) also support the use of behavioral assessment scales, because these scales also assess affective aspects of pain.

4.3. Clinical usefulness

Jensen (1997) stated that "no measure should be used without evidence for both its reliability and its validity". Although we agree, we want to emphasize that besides reliability and validity, usefulness is at least of equal importance. Therefore, from the start of this project, we highly valued nurses' opinions about the usefulness of the scales, as they are, after all, the main intended users. Analyses of clinical usefulness revealed that the nurses found PACSLAC to be the most useful and the preferred scale in this setting.

4.4. Limitations

Some limitations need to be acknowledged. First, we are aware that the methodology used in this study may have influenced findings. Elderly nursing home residents were observed in controlled situations with mostly a standardized acute pain stimulus. Raters were not blinded to the intervention, which may have affect the findings and their generalizability. Vaccinations were chosen as the standardized pain stimulus, as it is not easy to choose a pain stimulus which provides an opportunity to observe a considerable number of patients with a reasonable level of pain. We hypothesized that there would not be many extremely high pain scores in reaction to the injection. Our analyses confirmed that the injection caused low to moderate levels of pain in the majority of participants. However, we used additional pain moments at T3 to enhance the reliability of our results. Patients who were not expected to be in pain at any specific moment, according to the nurses, were not assessed a third time. Hence, nurses who observed these patients were well aware of the fact that pain might

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

be present. Another concern could relate to the fact that an acute pain moment was picked for this data collection. It is imaginable that more chronic pain moments would involve more frequent use of other items.

A second issue is our assumption that pain would be less prevalent at rest (T1) than at T2 and T3. Since elderly nursing home patients often suffer chronic pain, this assumption could be challenged.

4.5. Recommendations

The design of future studies should take account of knowledge we posses. Scherder et al. (2005) showed that the type of dementia probably plays an important role in pain experience. In the present study, the type of the dementia was often unknown as we depended on medical records for this information. Since we now know that the type of dementia does matter, this aspect should play a more prominent part in future studies.

The inability to diagnose pain adequately in nursing home residents with dementia is a significant problem for health care professionals. It is therefore an important step forward to have a clinically useful and psychometrically sound pain assessment scale available for this setting. The next step involves its implementation in nursing practice, which would allow further examination of psychometrics, like responsiveness or sensitivity to change. Future studies also need to focus on examining the prevalence of pain by using an observational pain scale.

This study took place in a highly controlled situation, in which elderly persons were observed in a standardized situation. In daily practice, the assessment of pain in elderly people with dementia is often influenced by many uncontrollable aspects, which add to the level of uncertainty. The challenge will be to examine the adequacy of scales in clinical practice and their effects on pain management.

Acknowledgements

This research was funded by the Research Council for Medical and Health Research of The Netherlands Organization for Scientific Research. The authors like to thank all the staff of the participating Dutch nursing homes (Hambos Kliniek in Kerkrade, Luckerheide Kliniek in Kerkrade, Grubbeveld in Maastricht and VKH in Heerlen) for their co-operation in this study.

References

Breau LM, McGrath PJ, Craig KD, Santor D, Cassidy K, Reid GJ. Facial expressions of children receiving immunizations: a principal components analysis of the Child Facial Coding System. Clin J Pain 2001;17:178–86.

Closs SJ, Barr B, Briggs M, Cash K, Seers K. A comparison of five pain assessment scales for nursing home residents with varying degrees of cognitive impairment. J Pain Symptom Manage 2004;27(3):196–205.

Craig KD, Prkachin KM, Grunau RE. The facial expression of pain. In: Melzack R, editor. Handbook of pain assessment. 2nd ed. New York, London: The Guilford Press; 2001. p. 153–69.

Farrell MJ, Katz B, Helme RD. The impact of dementia on the pain experience. Pain 1996;67:7–15.

Ferrell BA. Pain evaluation and management in the nursing home. Ann Intern Med 1995;123(9):681–7.

Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.

Fuchs S, Hadjistavropoulos H, McGrath P. Psychometric development of a pain assessment scale for older adults with severe dementia: a report on the first two studies. Paper presented at the IASP, San Diego; 2002.

Fuchs-Lacelle S, Hadjistavropoulos H. Development and preliminary validation of the Pain Assessment Checklist for Seniors with Limited Ability to Communicate (PACSLAC). Pain Manag Nurs 2004;5(1):37–49.

Hadjistavropoulos T, LaChapelle DL, MacLeod FK, Snider B, Craig KD. Measuring movement-exacerbated pain in cognitively impaired frail elders. Clin J Pain 2000;16:54–63.

Hartmaier SL, Sloane PD, Guess HA, Koch GG, Mitchell CM, Phillips CD. Validation of the minimum data set cognitive performance scale: agreement with the mini-mental state examination. J Gerontol 1995;50A(2):M128–33.

Helme RD, Gibson SJ. The epidemiology of pain in elderly people. Clin Geriatr Med 2001;17(3):417–31.

Hølen JC, Saltveldt I, Fayers PM, Bjørnnes M, Stenseth G, Hval B, et al. The Norwegian Doloplus-2, a tool for behavioural pain assessment: translation and pilot-validation in nursing home patients with cognitive impairment. Palliat Med 2005;19:411–7.

Hurley AC, Volicer BJ, Hanrahan PA, Houde S, Volicer L. Assessment of discomfort in advanced Alzheimer patients. Res Nurs Health 1992;15:369–77.

Jensen MP. Validity of self-report and observational measures. In: Jensen TS, Turner JA, Z Wiesenfeld-Hallin, editors. Proceedings of the 8th World Congress on Pain. Seattle: IASP Press, 1997. pp. 637–61.

Jensen MP, Karoly P. Self-report scales and procedures for assessing pain in adults. In: Melzack R, editor. Handbook of pain assessment. 2nd ed. New York, London: The Guilford Press; 2001. p. 14–34.

Leong IY, Chong MS, Gibson SJ. The use of a self-reported pain measure, a nurse-reported pain measure and the PAINAD in nursing home residents with moderate and severe dementia: a validation study. Age Ageing 2006;23:1–5.

Lane P, Kuntupis M, MacDonald S, McCarthy P, Panke JA, Warden V, et al. A pain assessment tool for people with advanced Alzheimer's and other progressive dementias. Home Healthcare Nurse 2003;21(1):32–7.

Lansbury G. Chronic pain management: a qualitative study of elderly people's preferred coping strategies and barriers to management. Disabil Rehabil 2000;22:2–14.

Lefebvre-Chapiro S, Doloplus group. The Doloplus 2 scale-evaluating pain in the elderly. Eur J Palliat Care 2001;8(5):191–4.

McGrath PA, Seifert CE, Speechley KN, Booth JC, Stitt L, Gibson MC. A new analogue scale for assessing children's pain: an initial validation study. Pain 1996;64(3):435–43.

McGrath PJ, Rosmus C, Canfield C, Campbell MA, Hennigar A. Behaviours caregivers use to determine pain in non-verbal, cognitively impaired individuals. Dev Med Child Neurol 1998;40:340–3.

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.

Michel M, Capriz F, Gentry A, Filbet M, Gauquelin S, Lefebre-Chapiro L, et al. Doloplus 2, une échelle comportementale de la douleur valide chez la personne âgée. La Revue de Gériatrie 2000;25(3):155–60.

Morris JN, Fries BE, Mehr DR, Hawes C, Phillips C, Mor V, et al. MDS cognitive performance scale. J Gerontol 1994;49(4):M174–82.

Nunnally Jr JC. Psychometric theory. 2nd ed. New York: McGraw-Hill; 1978.

Oberlander TF, Gilbert CA, Chambers CT, O'Donnell ME, Craig KD. Biobehavioral responses to acute pain in adolescents with a significant neurological impairment. Clin J Pain 1999;15:201–9.

Polit DF, Beck CT. Nursing research. Principles and methods. Philadelphia, PA: Lippincott, Williams & Wilkins; 2004.

Scherder E, Oosterman J, Swaab D, Herr K, Ooms M, Ribbe M, et al. Recent developments in pain in dementia. Br Med J 2005;26:330 (7489):461–464.

Scherder E, Slaets J, Deijen JB, Gorter Y, Ooms M, Ribbe M, et al. Pain assessment in patients with possible vascular dementia. Psychiatry 2003;66(2):133–45.

Serbouti S, Rat P, Passadori Y. Validation of DOLOPLUS. A pain assessment tool for non-verbal or cognitively impaired elderly patients. Poster presented at the IASP, San Diego; 2002.

Streiner DL, Norman CR. Health measurement scales. A practical guide to their development and use. Oxford: University Press; .

Warden V, Hurley AC, Volicer L. Development and psychometric evaluation of the Pain Assessment in Advanced Dementia (PAIN-AD) scale. J Am Med Dir Assoc 2003;4:9–15.

Wary B, Doloplus collectif. Doloplus-2, une échelle pour évaluer la douleur. Soins Gérontologie 1999;19:25–7.

Wary B, Serbouti S, Doloplus collectif. Validation d'une échelle d'évaluation comportementale de la douleur chez la personne âgée. Douleurs 2001;2(1):35–8.

Weissman DE, Matson S. Pain assessment and management in the long-term care setting. Theor Med 1999;20:31–43.

Zwakhalen SMG, Hamers JPH, Huijer Abu-Saad H, Berger MPF. Pain in elderly people with severe dementia: a systematic review of behavioural pain assessment tools. BMC Geriatr 2006;27(6):3.

Please cite this article as: Sandra M.G. Zwakhalen, et al., The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia, Pain (2006), doi:10.1016/j.pain.2006.06.029.