A psychometric evaluation of three pain rating scales for people with moderate to severe dementia

Running title: Evaluation of three pain scales for dementia

Authors:

Christine Neville, PhD RN RPN FACMHN

Associate Professor

Deputy Director, Ipswich Clinical School

The University of Queensland

School of Nursing and Midwifery

Remo Ostini, PhD

Research Fellow

The University of Queensland

Healthy Communities Research Centre

Correspondence:

Associate Professor Christine Neville

The University of Queensland

School of Nursing and Midwifery

Qld 4072, Australia

Email: christine.neville@uq.edu.au

Phone: 61 7 33811182

Fax: 61 7 33811166

Abstract

Background: Little comparative information exists regarding the reliability and validity of pain rating scales for nurses to assess pain in people with moderate to severe dementia in residential aged care facilities. The objective of this study was to evaluate the relative psychometric merits of the Abbey Pain Scale, the DOLOPLUS 2 Scale, and the Checklist of Non-verbal Pain Indicators Scale – three well-known pain rating scales that have previously been used to assess pain in nonverbal people with dementia. An observational study design was used. Nurses (n = 26) independently rated a cross-section of people with moderate to severe dementia (n = 126) on two occasions. The Abbey Pain Scale and the DOLOPLUS 2 Scale showed good psychometric qualities in terms of reliability and validity, including resistance to the influence of rater characteristics. The Checklist of Non-verbal Pain Indicators also had reasonable results but was not as psychometrically strong as the Abbey Pain Scale and DOLOPLUS 2 Scale. This study has provided comparative evidence for the reliability and validity of three pain rating scales in a single sample. These scales are strong, objective adjuncts in making comprehensive assessments of pain in people who are unable to self-report pain due to moderate to severe dementia, with each having their own strengths and weaknesses. The DOLOPLUS 2 Scale provides more reliable measurement, while the Abbey Pain Scale may be better suited than the other two scales for use by nurse raters who only occasionally use pain rating scales, or who have lower level nursing qualifications.

Key words: Abbey Pain Scale, Checklist of Non-verbal Pain Indicators, DOLOPLUS 2, Nursing home, Pain measurement, Pain assessment

Introduction

Despite a higher prevalence of chronic diseases such as arthritis and cancer, pain in older people and particularly those with dementia should not be dismissed as a part of normal ageing. For people with dementia living in residential aged care facilities (RACFs), pain occurs at a persistent rate. Prevalence of pain in this group has been recently reported at 43% (Leong, Chong & Gibson, 2006), 47% (Torvik et al., 2010) and 68% (Zwakhalen et al., 2009). The critical issue however is that people with moderate to severe dementia are especially at risk for unidentified and under-treated pain, despite there usually being no differences in the potential physical causes for pain (Neville, McCarthy & Laurent, 2006).

The adequate treatment of pain is also an area where nurses, who are primary carers and are responsible for the management of pain, are increasingly being held accountable (Abbey et al., 2004). The assessment of pain concentrates on the pain description; alleviating or aggravating factors; its impact on functional, psychological and social status; and nurses' observations of pain, often based on interpretations of non-verbal cues. Assessment of pain is an important component in the treatment of pain, and there is a need for manageable, valid and reliable tools to assess pain in people with dementia (Collett et al., 2007). A significant number of nurses are not aware of pain rating scales or do not use them routinely to justify their pain management interventions (McAuliffe et al., 2009, Neville, McCarthy & Laurent, 2006).

Standardized techniques of pain assessment are important for developing a credible care protocol. Pain rating scales help to systematize information gathering and eliminate some of the

difficulties for nurses in deciding when and what form of pain relief are needed (Abbey et al., 2004). Consolidating information on a narrower range of quality tools should increase the likelihood of the most suitable tool being chosen. A number of recent reviews have examined pain rating scales for people with severe dementia and identified some that are more appropriate than others although all scales reviewed were deemed to have significant limitations (Herr et al., 2010, van Herk et al., 2007, Zwakhalen et al., 2006). Limitations of current scales included that they have often been only tested by developers with very small sample sizes, tested in limited clinical settings and not with the nurses who have to use the scales and finally, they often lack reliability data. Zwakhalen et al. (2006) recommended that future research should not focus on developing more scales but on improving understanding of existing scales by further testing their validity and reliability. Scales must also be clinically useful in an environment that is often very busy and where staff working most closely with the person with dementia have only basic qualifications. Therefore scales should require minimal time to complete and be easy to understand and use. To meet these recommendations and requirements, three pain rating scales for people with moderate to severe dementia have been chosen for a more detailed examination in Australian RACFs. These are the Abbey Pain Scale (APS: Abbey et al., 2004), DOLOPLUS 2 (Lefebvre-Chapiro & the DOLOPLUS Group, 2001) and the Checklist of Non-Verbal Pain Indicators (CNPI: Feldt, 2000).

The APS has been recommended by The Australian Pain Society (2005) for the management of pain in residential aged care facilities. This scale has been used widely and although several psychometric aspects have been tested, there is no test-retest reliabilty data available and the scale has not been tested independently for its reliability and validity with people with moderate

to severe dementia in Australian RACFs. The DOLOPLUS 2 has widespread use in Europe and the United Kingdom (Holen et al 2005; NHS 2010; Pautex et al., 2006). It was originally developed in French and has since been translated into many other languages. The English version used in this study has had limited psychometric testing and none in Australia. The CNPI was developed in North America and has consistently rated well in many reviews (Herr, Bjoro & Decker, 2006, Herr et al., 2010, Zwakhalen et al., 2006) but has never been tested with an Australian population. The aim of this study was to conduct a psychometric evaluation of three pain rating scales for people with moderate to severe dementia who reside in Australian RACFs. The specific research questions are:

1. Are any of the scales more reliable when used with people with moderate to severe dementia in RACFs?
2. Are the scales valid for use with people with moderate to severe dementia in RACFs; particularly in terms of the factorial structure of the tests, and when used by a range of nurses working in RACF settings?

Methods

Participants and settings

An observational study design was used to answer these questions. The study involved pain rating scale administration by 26 nurses from four RACFs situated in south-east Queensland, Australia. Participants included 157 long-term residents of these same RACFs who all had a reported diagnosis of dementia in their clinical file. Comprehensive data are available for 126 residents from three of the RACFs and the results reported here are for these participants. Ethical approval was obtained from The University of Queensland's Behavioral and Social Sciences Ethical Review Committee. All participating nurses provided informed consent and proxy consent was obtained for the participants with dementia.

Instruments

Data collection took place using a questionnaire document that gathered demographic data about the nurses administering the questionnaire and of the residents with dementia who were the subjects of the pain assessments. Registered nurses at each RACF provided demographic data about themselves (gender, age, qualifications and length of time in dementia care) and about the RACF residents with dementia who were the rated for pain (gender, age, birthplace and current medications). Separately, the 26 nurses who administered the pain rating scales also provided demographic information about themselves (gender, age, qualifications and length of time in dementia care).

Three published scales were included for pain assessment. The Abbey Pain Scale (APS) is a unidimensional scale with items reflecting observable behaviors and physiological changes that are known indicators of pain (Abbey et al., 2004). Scale items were identified from well-known earlier studies and through rigorous consultation with international experts (Abbey et al., 2004, Hurley et al., 1992, Simons & Malabar, 1995). The APS measures the severity of pain with six items: vocalization, facial expression, change in body language, behavioral change, physiological change and physical changes. Each item has descriptive prompts to assist nurses with their observations and to enhance reliability. The nurse rates the current level of pain from 0 ('absent'), 1 ('mild'), 2 ('moderate') to 3 ('severe'). The scores are summed to give a total possible score of 18 with scores from 0-2 indicating no pain, 3-7 indicating mild pain, 8-13 indicating moderate pain and 14+ indicating severe pain. This pain score was established by a cross tabulating the new pain score against a holistic measure developed specifically for the pain scale development study. The APS has established face, content and concurrent validity. Published internal reliability is reasonably adequate ($\alpha = 0.74$) and despite some recognized testing limitations has shown modest correlations for the inter-rater reliability scores (Abbey et al., 2004). Abbey et al., (2004) were extremely mindful of the pressures within the environment in which this scale is to be used and ensured it could be used by a variety of RACF staff, taking less than one minute to complete.

The DOLOPLUS 2 (Lefebvre-Chapiro & the DOLOPLUS Group, 2001) was originally developed in French and has since been translated into other languages (Ando & Hishinuma, 2010, Holen, et al., 2005). The English version used in this study has had limited psychometric testing. The scale involves observations of patient behavior in ten different situations that could

potentially involve pain. Items include sleep, verbal reaction and behavioral symptoms. Ratings are made from 0 to 3 representing increasing intensity in pain. The score does not represent pain experience at a specific moment but reflects on the progression of experienced pain. For the French version - convergent validity with the Visual Analogue Scale-patient was significiant ($p<0.001$) and the DOLOPLUS 2 demonstrated good sensitivity. There was satisfactory stability on the re-test. A t-test analyzing the intra-observer differences found no significant differences for the total score or for item scores. An interrater correlation test between two physicians showed no significant difference ($p<0.001$) and good levels of internal consistency ($\alpha = 0.82$) were found (Zwakhalen et al., 2006). Limitations include a lack of information about the determination of cut-offs scores and the impairment level of the participants. The maximum score is 30, and a score of 5 represents pain, raising questions about the scales specificity (Zwakhalen et al., 2006). Questions also remain about how easy the items are to interpret with different levels of staff and the scale's clinical utility needs further testing at the bedside with large samples.

The Checklist of Non-Verbal Pain Indicators (CNPI: Feldt, 2000) is a behavioral observation scale at rest and during movement. Clustered items include restlessnes, rubbing and vocal complaints. An item is scored '1' if the behavior is observed and '0' if not observed (range of total scale 0-6). After adding up the two scores (for movement and rest), the interpretation is '1-2' for mild pain, '3-4' for moderate pain, '5-6' severe pain. Feldt (2000) claimed good face validity and good inter-rater reliability. Limitations include the CNPI only correlating with a Verbal Descriptive Scale during movement and these correlations were low ($r=.372$ at rest; $r=.428$ during movements). There were moderate levels of internal consistency ($\alpha = .54$ at rest; α

= .64 during movement) (Feldt, 2000). Based on reported findings, the CNPI has poor psychometric qualities and requires further testing (eg test-retest) in large samples (Zwakhalen et al., 2006).

An additional Yes/No question asked the registered nurse filling out the patient demographic section of the questionnaire whether the person with dementia had significant pain symptoms. This was included for the purpose of external comparison to determine the extent to which the pain assessment scales relate to the 'real' world practice of nurses relying on their clinical judgment to assess pain (Neville, McCarthy & Laurent, 2006; Parke, 1998).

Procedures

Table 1 outlines the data collection process. At the study outset, a registered nurse compiled demographic information for each RACF resident with dementia who was to be included in the study and completed the Yes/No question on the presence of significant pain. Thirty-two volunteer nurse participants across the four RACFs then received training from a project team member as raters of the pain scales. The participating nurse raters at each facility provided demographic data about themselves and then each participating resident had their pain rated by two independent nurse raters (assigned to rater group 1 or 2). Residents were rated for pain at a nominated time in mid-afternoon, with two testing occasions occuring two weeks apart. At the second testing occasion, participating residents had their pain independently rated by the same two nurse raters as on the first testing occasion. As recommended by Zwakhalen, Hamers and Berger (2006), the time interval between assessment scale administrations was kept under three weeks to minimise the risk of score changes due to actual changes in health status over time.

Analgesic administration during the study period was provided in a manner consistent with usual practice at each RACF, thus avoiding peaks and troughs of analgesia.

Table 1: Schedule of Measurements

Information Collected	Measurement Time	
	1	2
Demographic Questions (person with dementia)	Registered Nurse	
Demographic Questions (nurse)	Rater Groups 1& 2	
APS, DOLOPLUS 2, CNPI and other measures	Rater Groups 1 & 2	
APS, DOLOPLUS 2, CNPI		Rater Groups 1 & 2

Statistical analysis

Demographic characteristics for both nurse raters and participants with dementia were calculated using proportions for categorical variables (e.g. gender, nursing qualifications, dementia status, and medication use) together with the range, mean, and standard deviation for continuous

variables (e.g., age, and test scales, including pain scales). Mean scores were calculated for each pain scale, by rater group, at each time period. These mean scores were interpreted in terms of a descriptive label for the level of pain indicated by the means, as defined in the development of each pain scale.

A series of six multiple regression analyses were conducted to obtain an assessment of potential rater influences on pain scale scores. Independent variables included information about which nurse rater had produced a resident's pain score and rater demographics (age, qualifications, years working in dementia care). Pain scale scores at each testing occasion, for each pain scale, represented the dependent variables. A stepwise procedure was used, with demographic variables entered first, as a block, followed by the individual rater designation for each RACF resident. Applying a Bonferroni correction to account for inflated experiment-wise error rates resulted in the adoption of α=0.008 as the criterion for inferential test significance across the multiple regressions.

Subsequently, four analyses were conducted to investigate scale reliabilities. Pearson product-moment correlation coefficients were calculated to estimate test-retest reliability for each pain scale, by rater. Cronbachs' alpha coefficients were calculated for each pain scale on scores provided by each rater group at each time period, to obtain four estimates of internal consistency reliability for each pain scale. Intraclass correlation coefficients were calculated on the total scores of each pain scale to provide an estimate of inter-rater reliability for each pain scale at each of the two testing occasions. Total pain scores were then used to classify patients into pain levels (No pain, Mild pain, Moderate pain, Severe pain) using the recommended score ranges for

each scale. Weighted Kappa was calculated to compare these pain level categorizations across raters as a final measure of inter-rater reliability.

Two sets of factor analyses were conducted to investigate the construct validity of the three pain scales. First, item level exploratory factor analyses, with principal axis factoring, were conducted for each separate rater group-by-testing occasion condition (4 conditions for each pain scale). The eigenvalue greater than 1.0 heuristic was used to select the number of factors to extract in each condition. The unrotated results of these analyses were compared for adherence to simple structure with 1-factor solutions in each condition, in order to obtain an indication of the support for the unidimensionality assumption underlying the scoring of these pain scales.

A test score exploratory factor analysis, with principal axis factoring, was also conducted on the complete set of 12 test scores for each pain scale across each of the four (rater group by testing occasion) conditions. This analysis was conducted to see whether each pain scale's set scores (four for each scale) would load as a group, independently of the other pain scales' scores.

A final set of analyses was conducted to assess external relationships for the pain scales. Evidence of relationships for the three pain scales with the external pain assessment was obtained by calculating Pearson correlation coefficients between the initial registered nurse Yes/No rating of significant pain and the pain scale scores of each pain scale for each rater group at the first testing occasion.

Results

Characteristics of Nurses: All but one of the nurse raters participating in this study (N=26) were female and most were Australian born (60%). In this study, 58% were registered nurses, 33% were enrolled nurses (in Australia an enrolled nurse is the equivalent of a Licensed Practical Nurse), and 9% assistants-in-nursing (in Australia an assistant-in nursing is equivalent to a nurse's aide or a nurse technician). Forty-eight percent of the nurses were aged between 51 and 60 years. Forty-one percent of the nurses had been working in dementia care for between 11 and 20 years.

Characteristics of People with Dementia: Table 2 presents demographic and baseline characteristics for participating RACF residents with dementia. The majority of residents were female and over 80 years of age. All had a reported diagnosis of dementia with the majority experiencing moderately severe to very severe cognitive decline. This was a relatively dependent group of people for activity of daily living needs with a moderate level of chronic illness burden. Analgesics were the most commonly prescribed medication. On average the sample was depressed and exhibited a moderate to high frequency of behavioral and psychological symptoms of dementia (Table 2).

Table 2: Demographics of People with Dementia

Variable	*n*	%	*m* (sd)	Range
Persons	126			
Gender				
Male	22	17		
Female	104	83		
Age (in years)	121	96	85.2 (6.6)	69 – 96
Dementia	126	100		
CIRS-G (range 0 – 56)	121	96	9.1(3.9)	0 - 19
ADL (range 3 – 18)	120	95	14.09 (2.6)	7 - 18
GDS (range 0 – 7)	120	95	5.7 (1.5)	0 - 7
CSDD (range 0 – 40)	72	57	11.7 (6.3)	0 - 35
DBDS (range 0 – 112)	121	96	56.28 (15.45)	29 - 106
Medications				
Analgesics		71		
Cardiovascular		60		
Gastrointestinal		53		
Musculoskeletal		39		
Antidepressants		33		
Endocrine		30		
Anti-anxiety		21		
Anti-psychotic		19		
Genitourinary		18		

Respiratory	9
Immune	8

CIRS-G = Cumulative Illness Rating Scale for Geriatrics (Miller et al., 1992); ADL = Activities of Daily Living (Katz et al., 1963); GDS = Global Deterioration Scale (Reisberg et al., 1982); CSDD = Cornell Scale for Depression in Dementia (Alexopoulos et al., 1988); DBDS = Dementia Behavior Disturbance Scale (Baumgarten, Becker & Gauthier, 1990).

Pain Scale Means and Standard Deviations

Descriptive statistics for the three pain scales, by rater group (R1, R2) and testing session (T1, T2), are shown in Table 3. Across the four conditions the APS produces an average assessment of mild pain, the DOLOPLUS 2 produces an average assessment indicating the presence of pain, and the average CNPI assessment is of moderate pain.

Table 3: Pain Scale Means and Standard Deviations

		R1T1	R1T2	R2T1	R2T2
APS					
	Mean (SD)	3.68 (3.09)	2.76 (2.80)	3.31 (3.06)	2.97 (2.88)
DOLOPLUS 2					
	Mean (SD)	8.29 (6.52)	7.15 (5.97)	7.41 (6.29)	6.82 (5.87)
CNPI					
	Mean (SD)	3.23 (2.60)	2.74 (2.53)	3.15 (2.60)	3.07 (2.92)

Influence of Individual Raters and Rater Demographics

Rater demographics were significantly associated with pain scale scores at the first testing occasion for the DOLOPLUS 2 (R2=0.057, p=0.004) and the CNPI (R2=0.054.p=0.005). At the second testing occasion, rater demographics were only significantly associated with CNPI pain scale scores (R2=0.072, p=0.001). For all three significant models, Rater Qualifications were the only significant independent variable, indicating that more highly qualified nurse raters tended to assign higher ratings to RACF resident pain. In none of the six regression models did individual

rater designation add significantly to the explanatory power of the models (R^2_{change} ranged from 0.000 to 0.010, all $p>0.12$).

Reliability Analyses

Across test-retest, internal consistency and ICC reliability analyses (See Table 4), the DOLOPLUS 2 was the pain scale with the highest reliability coefficient; with the CNPI showing the best weighted Kappa results. The CNPI produced the lowest test-retest coefficients but outperformed the APS on internal consistency. Test-retest and inter-rater reliability coefficients were moderately good for the APS and the DOLOPLUS 2. Cronbach's alpha, internal consistency reliability coefficients for the pain scales were good (APS, CNPI) to very good (DOLOPLUS 2). The slightly higher alpha coefficients for the CNPI, when compared to the APS, were likely due to the CNPI being a much longer scale.

Item level internal consistency analysis consistently showed that the APS would be a more reliable scale if the item 'Physical Changes' were omitted from the scale. This same analysis showed that all DOLOPLUS 2 items contributed to the reliability of this scale. Results for the CNPI consistently showed that omitting the 'Rest Rubbing' item from the scale would result in a small improvement in the scale's internal consistency reliability.

Table 4: Reliability of APS, DOLOPLUS 2, and CNPI

Test-Retest	Rater 1		Rater 2	
APS	0.680		0.618	
DOLOPLUS 2	0.707		0.706	
CNPI	0.564		0.443	
Internal Consistency	**Rater 1**		**Rater 2**	
	Time 1	Time 2	Time 1	Time 2
APS	0.736	0.796	0.650	0.779
DOLOPLUS 2	0.857	0.864	0.867	0.864
CNPI	0.756	0.765	0.785	0.822
Inter-rater (ICC)	**Time 1**		**Time 2**	
APS	0.750		0.704	
DOLOPLUS 2	0.733		0.812	
CNPI	0.586		0.713	
Inter-rater (weighted Kappa)	**Time 1**		**Time 2**	
APS	0.335		0.475	
DOLOPLUS 2	0.421		0.501	

CNPI	0.432	0.532

ICC- Intraclass correlation coefficient

All test-retest and inter-rater coefficients significant at $p<0.001$

Factor Analyses

Factor analyses were conducted to determine whether the pain scales could best be described with a 1-factor model, or whether a more complex model was needed. Factor analyses, using principle axis factoring, were conducted separately on data from each pain scale, in each of the four test administration settings (two rater groups by two testing occasions).

The size of the sample was not large, for factor analysis purposes, and this may have contributed to some variation in results across analyses. Nevertheless, a clear pattern of results was usually evident across the analyses for each pain scale.

Factor analysis of the APS data produced a set of results indicating that a 1-factor solution was the best solution for describing the psychometric structure of this scale. It was also found that the item 'Physical Changes' often did not load strongly on this factor. A single-factor model typically accounted for approximately 40% of the scale variance as common variance.

Factor analysis of the DOLOPLUS 2 data similarly found that a 1-factor solution was the best description of the factor structure of this scale. This solution also typically accounted for approximately 40% of the available variance as common variance.

Factor analysis results for the CNPI were less clear-cut than for the two other pain scales. In the first instance, the 1-factor solutions typically only accounted for 25-30% of the available variance as common variance. However, it was not as clear as for the other scales, that a 1-factor solution was the best solution for this pain scale. There was also some support for a 2-factor solution. This solution essentially broke down into a 'Rest' factor and a 'Movement' factor. It was also found that the 'Rubbing' items on the CNPI (both Rest and Movement) failed to load on factors that were extracted. The more equivocal results for the CNPI may be influenced by the relatively modest sample size for this, the longest pain scale.

A final, test level, factor analysis was conducted on the correlations among the 12 total scale scores for the pain scales (3 scales by 2 rater groups by 2 testing occasions). The intent of this analysis was to see whether any of the scale's total scores, or more interestingly, the set of scores for any scale, rater or occasion, were less closely related to the remaining scores. This did not occur. The factor analysis showed that a single factor model best described the correlations among all 12 total scale scores with each score loading highly (>0.60) on that single factor. This indicates that all of the scales, as administered by both rater groups on each occasion measured essentially the same single construct.

External Validity

Each of the scales showed medium correlations (Cohen, 1988, pp. 79-80) with the Yes/No response initially completed by a registered nurse for all participating residents at each RACF, independently of the pain scale nurse raters, to the external validity question of *Does the older adult have significant pain symptoms?* (59% of older adults were reported to have significant

pain symptoms of a chronic nature). In all cases these results were significant at α=0.01 (See Table 5).

Table 5: Pearson product-moment correlation coefficients between APS, DOLOPLUS 2, CNPI, pain scores at first testing occasion and independent registered nurse pain rating

	Initial Independent Pain Rating
Rater Group 1	
APS	.38**
DOLOPLUS 2	.43**
CNPI	.34**
Rater Group 2	
APS	.45**
DOLOPLUS 2	.45**
CNPI	.40**

** $p < .01$ (2-tailed)

Discussion

The findings of this study provide evidence on the relative psychometric strengths of three pain rating scales (APS, DOLOPLUS 2 and the CNPI) for people with moderate to severe dementia residing in Australian RACFs. These results indicate that all three scales were measuring a similar conception of pain.

The APS was found to have good validity and reliability, although it could be refined further. If the item 'Physical Changes' were dropped, the homogeneity of the scale would be enhanced. The DOLOPLUS 2 showed the strongest results across three of the four reliability analyses, as well as showing good homogeneity and validity. The CNPI demonstrated the strongest inter-rater reliability results and adequate levels of internal consistency (except for the 'Rubbing' item at rest and movement). Closer examination revealed that it is measuring pain at rest and at movement but the process of combining the scores is questionable. This scale also demonstrated the most susceptibility to the effect of different nurse rater qualification levels.

Influence of Raters and Rater Demographics

Neither the different raters, nor the demographic characteristics of those nurse raters had a significant impact on APS scores on either testing occasion. Indeed, there was no significant effect from 26 different nurse raters producing pain ratings for 126 RACF residents, over and above the effects of rater demographics, for any of the pain scales. This result speaks positively

to the validity of the pain scores produced by the data collection process used in this study, indicating that the multiple rater process did not bias pain scores.

The demographic characteristic of the raters did, however, have a significant effect on the DOLOPLUS 2 and the CNPI pain scores at the first rating occasion; and again on the CNPI scores at the second rating occasion. In each case, the only significant individual predictor of pain scores was nurse qualifications. This set of results suggests that the APS is most impervious to the effect of different rater characteristics; the DOLOPLUS 2 is initially susceptible to rater characteristics but this effect dissipates with repeated use of the scale; while the CNPI is most susceptible to the effect of rater characteristics, specifically nurse qualifications, and that this effect is not ameliorated by rater familiarity with the scale.

The direction of the results suggest that initially with the DOLOPLUS 2 and on both occasions with the CNPI, less qualified nurse raters were under-evaluating resident's pain.

Reliability and Item Factor Analysis

The results of this study showed adequate levels of internal consistency for the APS, although 'Physical Changes' was the poorest scoring item on the APS. The APS may perform more satisfactorily in a population who experience predominantly chronic pain with this item deleted. Physiological changes such as increased breathing, noisy labored breathing, increased heart rates, perspiration, and flushing is observed may be more descriptive of acute pain as opposed to chronic pain (Lui, Briggs & Closs, 2010). Test-retest and inter-rater reliability coefficients were

moderately good for the APS and factor analysis supported a single factor structure – particularly if 'Physical Changes' is omitted.

The Chronbach's alpha was very good for the total DOLOPLUS 2 scale at Time 1 and Time 2 and was higher than the result reported by a Norwegian nursing home study (α=0.71) with people with severe cognitive impairment (Torvik et al., 2010). The factor analysis showed that a single factor described the DOLOPLUS 2 well. Test-retest and inter-rater reliability coefficients were moderately good for the DOLOPLUS 2. These findings are supported by a study that examined the psychometric qualities of the DOLOPLUS 2 with 128 people with dementia in Dutch nursing homes (Zwakhalen et al., 2006).

Most of the psychometric findings were weaker for the CNPI in measuring pain in people with severe to moderate dementia. The notable exception was the CNPI's stronger inter-rater agreement results. The findings of the factor analysis showed that the CNPI might consist of more than one, single factor. This should not be taken as evidence to conclude that the CNPI is inferior. In this study we used a purposive sample of aged care facility residents with dementia and therefore generalization of the findings in the present study is limited to similar populations. The original measured construct of the CNPI (pain in older people with hip fracture on a hospital surgical ward) is different from the construct (that is, chronic pain) in this study.

Additional Validity Indicators

The fact that the three pain scales' total scores load on a single factor and correlate with an external pain rating bodes well for their current and probable continued use by nurses caring for people with moderate to severe dementia. The finding also supports the conclusion that the best pain estimate is obtained by use of a systematic approach and trained nurses who are well aware of the person they are caring for (Torvik et al., 2010).

Limitations

Some limitations are to be noted. RACF residents with dementia were rated at rest with consideration given as to what had been their experience of routine nursing care on that day (for example, being moved from their bed to a chair). However, there are aspects of the RACF environment that could not be controlled, which leaves the results open to the effects of confounding factors but supports the ecological validity of the results. Similarly, the use of a group of nurse raters increases the potential for measurement error but also means that the results reflect the use of these scales under real world conditions.

Further validation studies of pain rating scales are needed among people where different types of dementia are specified, as the type of dementia may matter to the pain ratings produced. Studies to examine the responsiveness and sensitivity to change of the pain scale assessments are also required.

Conclusions

The CNPI appears less well-suited for measuring chronic pain in RACF residents with dimentia than the APS or the DOLOPLUS 2. The pattern of reliability results was weaker for the CNPI and it also had the weakest validity results, both in terms of its factor structure and its susceptibility to the influence of rater differences. THE APS and the DOLOPLUS 2 appear better suited to measuring pain in people with the characteristics of this study sample. Both measures were well-supported by the factor analysis results. The DOLOPLUS 2 showed stronger reliability indicators but was also initially more susceptible to the influence of rater characteristics.

The lack of a suitable pain rating scale can mean inadequate information for maximizing a person with dementia's well-being; or frustration for nurses who are unable to effectively assess pain. For nurses to correctly quantify pain in people with dementia they need valid and reliable rating scales. These scales should supplement clinical judgment and provide a standardized method to communicate and document pain. The results of this study suggest that the the DOLOPLUS 2 and the APS are more likely to meet this need for the benefit of nurses and people with moderate to severe dementia in settings like Australian RACFs.

Acknowledgements

The authors would like to acknowledge the generous support given to us by the residents and staff of the participating residential aged care facilities. Additionally, the funding provided by the Early Career Researcher Scheme at The University of Queensland.

References

Abbey, J., Piller, N., De Bellis, A., Esterman, A., Parker, D., Giles, L., & Lowcay, B. (2004). The abbey pain scale: A 1-minute numerical indicator for people with end stage dementia. *International Journal of Palliative Nursing,* 10(1), 6-14. Retrieved from http://www.internurse.com.ezproxy.library.uq.edu.au/cgi-bin/go.pl/library/article.cgi?uid=12013;article=IJPN_10_1_6_13

Alexopoulos, G.S., Abrams, R.C., Young, R.C., & Shamoian, C.A. (1988). Cornell scale for depression in dementia. *Biological Psychiatry,* 23(3), 271-284. doi.org.ezproxy.library.uq.edu.au/10.1016/0006-3223(88)90038-8

Ando, C., & Hishinuma, M. (2010). Development of the Japanese DOLOPLUS-2: A pain assessment scale for the elderly with alzheimer's disease. *Psychogeriatrics,* 10(3)**,** 131-137. doi: 10.1111/j.1479-8301.2010.00324

Baumgarten, M., Becker, R., & Gauthier, S. (1990). Validity and reliability of the dementia behavior disturbance scale. *Journal of the American Geriatrics Society,* 38(3), 221-226.

Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences*. Hillside: Lawrence Erlbaum Associates.

Collett, B., O'Mahoney, S., Schofield, P., Closs, S.J., & Potter, J. (2007). The assessment of pain in older people. *Clinical Medicine,* 7(5), 496-500. Retrieved from http://www.clinmed.rcpjournal.org/content/7/5/496.long

Feldt, K. (2000). The checklist of nonverbal pain indicators. *Pain Management Nursing,* 1(1), 13-21. doi.org.ezproxy.library.uq.edu.au/10.1053/jpmn.2000.5831

Herr, K., Bjoro, K., & Decker, S. (2006). Tools for assessment of pain in nonverbal older Adults with dementia: A state-of-the-science review. *Journal of Pain and Symptom Management,* 31(2), 170-192. doi: 10.1016/j.jpainsymman.2005.07.001

Herr, K., Bursch, H., Miller, L.L., & Swafford, K. (2010). Use of pain-behavioral assessment tools in the nursing home: Expert consensus recommendations for practice. *Journal of Gerontological Nursing,* 36(3), 18-39. doi: 10.3928/00989134-20100108-04

Holen, J.C., Saltvedt, I., Fayers, P.M., Bjornnes, M., Stenseth, G., Hval, B., Filbet, M., Loge, J.H., & Kaasa, S. (2005). The Norwegian doloplus-2, a tool for behavioural pain assessment: Translation and pilot-validation in nursing home patients with cognitive impairment. *Palliative Medicine,* 19(5), 411-417. doi: 10.1191/0269216305pm1031oa

Hurley, A., Volicer, B., Hanrahan, P., Houde, S., & Volicer, I. (1992). Assessment of discomfort in advanced alzheimer patients. *Research in Nursing and Health,* 15(5), 369-377. doi: 10.1002/nur.4770150506

Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., & Jaffe, M.W. (1963). Studies of illness in the aged. The index of ADL: A standardized measure of biological and psychosocial function. *JAMA*, 185(12), 914-919. doi:10.1001/jama.1963.03060120024016

Lefebvre-Chapiro, S., & DOLOPLUS Group. (2001). The DOLOPLUS® 2 scale – evaluating pain in the elderly. *European Journal of Palliative Care*, 8, 191-194.

Leong, I.Y.O., Chong, M.S., & Gibson, S. (2006). 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*, 35(3), 252-256. doi: 10.1093/ageing/afj058

Lui, J.Y.W., Briggs, M., & Closs, S.J. (2010). The psychometric qualities of four observational pain tools (OPTs) for the assessment of pain in elderly people with osteoarthritic pain. *Journal of Pain and Symptom Management*, 40(4), 582-598. doi: 10.1016/j.jpainsymman.2010.02.022.

McAuliffe, L., Nay, R., O'Donnell, M., & Fetherstonhaugh, D. (2009). Pain assessment in people with dementia: Literature review. *Journal of Advanced Nursing*, 65(1), 2-10. doi: 10.1111/j.1365-2648.2008.04861.x.

Miller, M.D., Paradis, C.F., Houck, P.R., Mazumder, S., Stack, J.A., Hind Rifai, A., Mulsant, B., & Reynolds III, C.F. (1992). Rating chronic medical illness burden in geropsychiatric practice

and research: Application of the cumulative illness rating scale. *Psychiatry Research*, 41(3), 237-248. doi.org/10.1016/0165-1781(92)90005-N

Neville, C., McCarthy, A., & Laurent, K. (2006). Pain management skills of regional nurses caring for older people with dementia: A needs analysis. *Collegian*, 13(2), 31-36. doi.org.ezproxy.library.uq.edu.au/10.1016/S1322-7696(08)60522-3

NHS *Lanarkshire Dementia Resource Folder*. (2010). Retrieved from http://www.knowledge.scot.nhs.uk/media/CLT/ResourceUploads/1010267/Dementia%20resource%20Folder%20Feb%202010.pdf accessed 09/28/2013.

Parke, B. (1998). Gerontological nurses' ways of knowing. Realizing the presence of pain in cognitively impaired older adults. *Journal of Gerontological Nursing*, 24(6), 21-28.

Pautex, S., Michon, A., Guedira, M., Emond, H., Le Lous, P., Samaras, D., Michel, J.P., Herrmann, F., Giannakopoulos, P., & Gold, G. (2006). Pain in severe dementia: Self assessment or observational scales? *Journal of the American Geriatrics Society*, 54(7), 1040-1045. doi :10.1111/j.1532-5415.2006.00766.x

Reisberg, B., Ferris, S.H., de Leon, M.J., & Crook, T. (1982). The global deterioration scale for assessment of primary degenerative dementia. *American Journal of Psychiatry*, 139(9), 1136-1139.

Simons, W., & Malabar, R. (1995). Assessing pain in elderly patients who cannot respond verbally. *Journal of Advanced Nursing*, 22(4), 663-669. doi: 10.1046/j.1365-2648.1995.22040663.x

The Australian Pain Society. (2005). *Pain in Residential Aged Care Facilities: Management Strategies*. North Sydney: The Australian Pain Society.

Torvik, K., Stein, K., Kirkevold, O., Saltvedt, I., Holen, J.C., Fayers, P., & Rustoen, T. (2010). Validation of Doloplus-2 among nonverbal nursing home patients – an evaluation of Doloplus-2 in a clinical setting. *BMC Geriatrics,* 10, 9. doi: 10.1186/1471-2318-10-9

van Herk, R., van Dijk, M., Barr, F.P.M., Tibboel, D., & de Wit, R. (2007). Observation scales for pain assessment in older adults with cognitive impairments or communication difficulties. *Nursing Research,* 56(1), 34-42. Retrieved from http://cf5pm8sz2l.search.serialssolutions.com/OpenURL_local?sid=Entrez:PubMed&id=pmid:17179872

Zwakhalen, S.M.G., Hamers, J.P.H., Huijer Abu-Saad, H., & Berger, M.P.F. (2006). Pain in elderly people with severe dementia: A systematic review of behavioural pain assessment tools. *BMC Geriatrics,* 27, 3. doi: 10.1186/1471-2318-6-3

Zwakhalen, S.M.G., Hamers, J.P.H., & Berger, M.P.F. (2006). The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia. *Pain,* 126(1-3), 210-220. doi.org.ezproxy.library.uq.edu.au/10.1016/j.pain.2006.06.029

Zwakhalen, S.M.G., Koopmans, R.T.C.M., Geels, P.J.E.M., Berger, M.P.F., & Hamers, J.P.H. (2009). The prevalence of pain in nursing home residents with dementia measured using an observational pain scale. *European Journal of Pain,* 13(1), 89-93. doi: 10.1016/j.ejpain.2008.02.009