JNM

Journal of Nursing Measurement

SPRINGER PUBLISHING COMPANY

www.springerpub.com/jnm

Journal of Nursing Measurement, Volume 25, Number 3, 2017

The Inter-Rater Reliability of the Observation Instrument for Assessing Pain in Elderly With Dementia: An Investigation in the Long-Term Care Setting

Gerhard Mueller, Assoc.-Prof. Dr., RN
Petra Schumacher, MScN, RN

UMIT–Private University for Health Sciences, Medical Informatics and Technology, Tyrol, Austria

Elisabeth Holzer, MScN

State Hospital, Lienz, Austria

Martin Pallauf, Dr.
Eva Schulc, Priv.-Doz. Dr.

UMIT–Private University for Health Sciences, Medical Informatics and Technology, Tyrol, Austria

Background and Purpose: The Observation Instrument for Assessing Pain in Elderly With Dementia (BISAD) was developed in Germany. The instrument demonstrated high inter-rater reliability values for the original French version. So far, there are no results to that effect in the Austrian long-term care setting available. The objective of this study was to investigate agreement and inter-rater reliability of BISAD in residents with dementia. Methods: A quantitative multicenter-descriptive cross-sectional design with a convenience sample of 71 residents. Results: Analysis of all eight items demonstrated fair to moderate concordance. Absolute agreement of the total value was 25.32%. Subtotals of the observation before mobilization was 52.11%, and during mobilization 32.39%. Conclusion: The reliability analysis shows that the items are less reliable. Currently, BISAD does not make a reliable contribution to clinical decision making in the tested setting.

Keywords: Observation Instrument for Assessing Pain in Elderly with Dementia (BISAD); pain measurement; dementia; inter-rater reliability; nursing home residents

The Berlin Institute for Population and Development (2011) describes dementia as the most common disease older than the age of 65 years. Worldwide, about 35.6 million people live with dementia (World Health Organization, 2012). In Germany, 1,572,104 people lived with dementia in 2013, which accounts to 1.92% of the total population. At the same time, dementia affected 145,432 people in Austria, which accounts to 1.73% of the total population (Alzheimer Europe, 2013). Furthermore, the European average of older adults with dementia was 1.55% of the total population; however, the disease is not diagnosed in 50%–60% of those being affected.

http://dx.doi.org/10.1891/1061-3749.25.3.E173

People in long-term care facilities usually suffer from multimorbidity, which may cause pain. Different causes of pain are always possible. Existing data on pain prevalence in nursing homes range between 12% and 68% (Centmayer & Lahmann, 2014; Mayer, Nonn, Osterbrink, & Evers, 2004; Osterbrink et al., 2012). Pain in dementia is related to comorbid conditions (Osterbrink et al., 2012), and pain recognition is a major challenge for nurses because residents with severe dementia cannot verbalize their pain adequately (Fischer, 2012; Nestler, Gnass, & Schuler, 2015; Schreier, Stering, Pitzer, Iglseder, & Osterbrink, 2015). As a result, the loss of the ability to communicate leads to the risk of insufficient pain recognition, assessment and treatment (Fischer, 2012; Osterbrink et al., 2012; Thomm, 2011).

BACKGROUND AND CONCEPTUAL FRAMEWORK

As occupational group with the most frequent contacts with residents, nurses have the opportunity to identify behaviors that cause pain (American Geriatric Society, 2002). Different conceptual models emphasize the role of external pain assessment by assessing the behavior in addition to self-assessment (Hadjistavropoulos & Craig, 2002; Snow et al., 2004). Thus, reliable pain assessment in persons with dementia is a complex task (Fischer, 2012) which, on the one hand, affects the pain diagnosis (Thomm, 2011) and, on the other hand, the pain therapy (Arnold et al., 2014).

Successful pain management depends, among other things, on the used pain assessment instrument. Internationally, 24 instruments exist for the pain assessment in nonverbally communicating (NVC) older adults, some developed especially for people with dementia (Fischer, 2012; Lichtner et al., 2014). German-speaking long-term care facilities use the German version of the Pain Assessment in Advanced Dementia (PAINAD) scale (Beurteilung von Schmerzen bei Demenz—BESD) and the Observation Instrument for Assessing Pain in Elderly With Dementia (Beobachtungsinstrument für das Schmerzassessment bei alten Menschen mit Demenz—BISAD), even though only first validation approaches have been undertaken (Fischer, 2012; Osterbrink et al., 2012; Thomm, 2011). BISAD was translated, developed, and validated by Fischer (2012) for assessing pain in people with moderate and severe dementia in 27 long-term care facilities in Germany. The scale is now used, although the original French version of BISAD was developed and tested on NVC older patients with different pain etiologies, including dementia (Morello, Jean, Alix, Sellin-Peres, & Fermanian, 2007). So far, no results on inter-rater reliability are available for the long-term care setting in Austria. Therefore, the aim of this study was to investigate the agreement and inter-rater reliability of BISAD (Fischer, 2012) in nursing home residents with dementia who had a Mini-Mental State Examination (MMSE; Kessler, Markowitsch, & Denzler, 2000) of $\leq$20 points in the long-term care setting.

DEVELOPMENT, ADMINISTRATION, AND SCORING OF OBSERVATION INSTRUMENT FOR ASSESSING PAIN IN ELDERLY WITH DEMENTIA

The instrument Echelle comportemental de la douleur pur personnes âgées non communicantes (ECPA) or Elderly Pain Caring Assessment 2 (EPCA-2; $n = 8$ items) was developed in France by Morello et al. (2007). It incorporates common pain behaviors demonstrated in NVC older patients (Fischer, 2012; Snow & Shuster, 2006) described by the American

Geriatric Society (2002). The scale was developed in several stages. In Stage 1, two geriatric experts on pain management conducted a literature review to research common pain behaviors in healthy persons and patients (Morello et al., 2007). In addition, 48 nursing and caregiving experts were asked to describe pain behaviors they had experienced in NVC older adults. During the review process of the literature and survey results, the two dimensions (before and during caregiving situations) of the scale were developed and the pain behavioral signs condensed to an 11-item first version of ECPA. In Stage 2, 66 NVC patients were rated with the 11-item ECPA version by a doctor pair ($n = 2$) and a nurse pair ($n = 2$) to establish face validity, inter-rater reliability, internal consistency, and tested factor analysis. Based on the results, one item was deleted and two items linguistically adapted. In Stage 3, the 10-item version was tested on 72 NVC patients with the same procedure as in Stage 2. One item was deleted because it did not load in one of the two dimensions of the scale. The third version of the ECPA (nine items) was tested on 78 patients following the same procedure as in the previous stages. After the review of the results, one item was deleted to establish internal consistency ($\alpha \leq .70$), thus leading to the final version of EPCA-2 with eight items (Morello et al., 2007). The EPCA-2 was tested for its psychometric properties on 340 NVC older patients with different pain etiologies, including 68 patients with dementia by different rater pairs (nurses, caregivers, doctors) in three university hospital centers in France. The inter-rater reliability verification presented very high agreement (intraclass correlation coefficient [ICC] = .877, 95% confidence interval or CI [.85, .89]). The internal consistency was $\alpha = .79$. The face and content validity were confirmed as good by all raters and the convergent validity established with a global clinical score (GCS) ranging from 0 to 10 points (0 = no pain, 10 = severe pain) by expert raters (nurses and doctors with several years of experience in the pain assessment of NVC patients; $r = .864$, 95% CI [.813, .874]) as well as the discriminant validity ("change in EPCA-2 scores" and the "change in GCS," $r_s = .619$, 95% CI [.490, .722] and divergent validity (patient's age and his or her EPCA-2 score, $r_s = .020$, 95% CI [−.087, .126] (Morello et al., 2007).

For the BISAD development, the EPCA-2 was translated with permission from the developers (Morello et al., 2007) according to the recommended translation process by Hilton and Skrutkowski (2002) by three experts into German and then translated back by two French speakers (Fischer, 2012). During this process, three German items had to be adapted for semantic equivalence. BISAD differentiates between two observation times (Areas 1 and 2). Area 1 contains four items that are observed before mobilization (e.g., sitting in a chair or lying in bed). The four items in Area 2 are observed during mobilization (e.g., walking, transfers in and out of bed) because pain may be intensified by movement (Gibson, 2006). All items can be rated on a 5-point Likert scale (ranging from 0 to 4; Table 1). A subtotal is determined for both observation times. The total value (subtotals of Areas 1 and 2) range between 0 and 32 points (Fischer, 2012). After the completion of the translation process, BISAD was tested for its validity and reliability in elderly nursing home residents with moderate ($n = 81$) and severe dementia ($n = 68$) by their key caregiver. First, the 149 residents were asked to rate their pain level with the verbal rating scale. Only 23 persons were able to rate their pain, even though residents with moderate dementia were still able to verbally communicate (Fischer, 2012). The results of the first validation study by Fischer (2012) presents moderate internal consistency of BISAD (Area 1: $\alpha = .64$, Area 2: $\alpha = .65$). The construct validity was tested by factor analysis with principal component analysis, and the varimax rotation explained 63% of variance. In addition, Fischer (2012) demonstrated feasibility and user-friendliness of BISAD.

TABLE 1. Items of the Observation Instrument for Assessing Pain in Elderly With Dementia

Items	Area 1: Observation Before Mobilization (5-Point Likert Scale)
1	Facial expression: gaze and mimic [relaxed look (0), totally rigid expression (4)]
2	Spontaneous resting position [no relieving posture (0); immobile, paralyzed by pain (4)]
3	Movement (or mobilization) of person (in or out of bed) [moving/not moving as usual (0); immobility, great agitation (4)]
4	Interaction to others [makes contact as usual (0), complete indifference (4)]
Items	**Area 2: Observation During Mobilization (5-Point Likert Scale)**
5	Anxious anticipation during caregiving [no anxiety (0); screams, groans, moans (4)]
6	Reactions during mobilization [can be mobilized/moves alone (0), shows resistance to being moved (4)]
7	Reactions during caregiving of painful areas [no reactions (0), painful areas cannot be approached (4)]
8	Complaints during caregiving [does not complain (0), shouts and complains violently (4)]

Note. Observation Instrument for Assessing Pain in Elderly With Dementia is not available in English. The items were translated by the authors for the readers' clarification.

METHODS

Design and Participants

The study applied a quantitative multicenter-descriptive cross-sectional design. Because of the exploratory study process, sample size determination was not done. A convenience sample of 71 nursing home residents were selected in three Austrian nursing homes. All registered nurses in the participating nursing homes were asked to take part in the study. Regarding ethical considerations, the study was approved by a local Research Committee for Scientific and Ethical Questions (RCSEQ, Nr.1113/14). The nursing sample consisted of 46 registered nurses. Nurses could participate in the study if they were registered nurses with at least 2 years of professional experience to have gained competence in performing assessments and nursing care plans (Benner, 2004). In addition, the nurses had a full-time position, were entrusted with the resident's nursing care and support for two weeks, had a standardized training on the use of BISAD (Fischer, 2012) and MMSE (Kessler et al., 2000), and gave informed consent. The residents were selected by a registered nurse on shift who conducted an informed consent discussion with the resident or the resident's legally accepted representative. The nurse also verified the inclusion and exclusion criteria

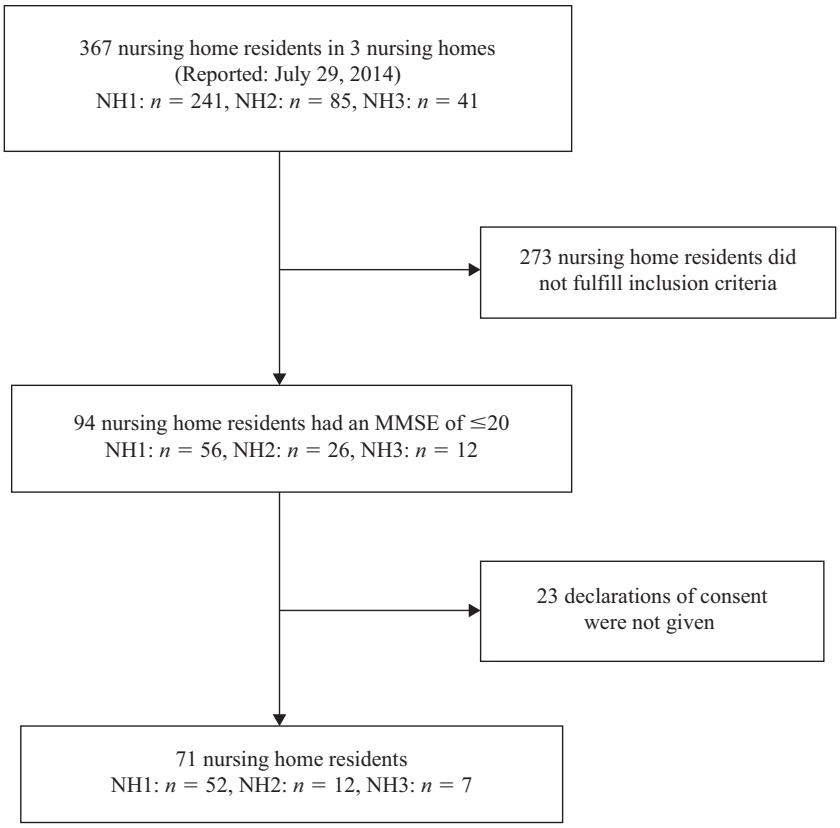

Figure 1. Sampling of nursing home residents. NH = nursing homes; MMSE = Mini-Mental State Examination.

(inclusion criteria: diagnosed dementia, an MMSE score of below or equal 20 points, permanent residency in the nursing home, age older than 65 years; exclusion criteria: MMSE with negative test result or 0 points; Figure 1).

Data Collection

This study was conducted in spring 2015. After written informed consent was given, a trained nurse on shift assessed the memory performance of the elderly resident with diagnosed dementia. Progression of cognitive impairment and severity of dementia was recorded with the German version of the MMSE (Kessler et al., 2000). Depending on the result, the resident was included or excluded from participating in the study. The test lasts a maximum of 10 min, depending on the cognitive impairment of the person (Buehler, 2014; Hoefler, Bengough, Winkler, & Griebler, 2015). The severity can be divided into three categories: severe (MMSE 0–11 points), moderate (MMSE 12–20 points), and mild dementia (MMSE 21–26 points; Hoefler et al., 2015; Kessler et al., 2000). MMSE demonstrated satisfactory validity and reliability values (Ahlsdorf & Schroeder, 2011).

To assess an instrument's reliability and precision (inter-rater reliability), Wirtz and Caspar (2002) recommend the assessment by two independent raters. The included residents were assessed independently by two raters (registered nurses on shift) with the

BISAD. The rater pairs were not always the same two nurses because they sometimes had different work schedules. The raters completed the resident's routine nursing care and independently assessed pain with BISAD afterward. This was completed once in a 24-hr period based on their observations during routine patient care. In addition, nurses were informed to conduct the assessment within 1 hr of each other, and an information exchange was not permitted between the two raters about the BISAD results.

The participating registered nurses completed a 1.5-hr comprehensive, standardized training, including specific information and application of BISAD and MMSE as well as the study's objectives and reasons. In addition, prior to the actual investigation, a pretest was carried out to verify the effectiveness of the training. For the pretest, three residents with dementia were rated with the MMSE and BISAD by two rater pairs ($n = 4$). The assessment duration lasted between 3 and 8 min for each rater. In a feedback discussion about their perceived training and assessment process, all nurses indicated the training as comprehensible and no problems were identified administrating MMSE or assessing pain with BISAD. The pretest results were not included in the final data analysis.

Statistical Analysis

Data analysis was performed by using SPSS Version 22. The extend of agreement was measured on item level, for the subtotals (Areas 1 and 2) as well as for the total value between a rater pair (two registered nurses) with absolute agreement (P_a), chance-corrected agreement with weighted Kappa coefficient (κ_w), and $ICC_{1.1}$ with 95% CI (Gwet, 2012; Wirtz & Caspar, 2002). The $ICC_{1.1}$ (one-way random single measures) was used because not all residents were assessed by the same rater pair (Wirtz & Caspar, 2002). According to Wirtz and Caspar (2002), the ICC can be calculated with ordinal data if a suitable reliability coefficient is calculated on an ordinal scale of measurement. If the interval data as well as the ordinal data indicate satisfying results, the question of the scale of measure is not critical for the assumption of data reliability. A minimum of 70% P_a was specified (Stemler, 2004). Reliability coefficients (κ_w, $ICC_{1.1}$) were interpreted with the standard conventions by Landis and Koch (1977). A possible difference between Area 1 (before mobilization) and Area 2 (during mobilization) was tested with the Wilcoxon test (Field, 2013). The significance level was set at 5% ($\alpha \leq .05$).

RESULTS

Participants

Seventy-one residents were included in the study; 84.5% of those were women. The participants were between 75 and 101 years old ($M = 89.55$, $SD \pm 5.85$). The average duration of stay in the nursing home was 5.42 ($SD \pm 4.28$) years. The residents' mean MMSE score as indicator for progression of cognitive impairment and severity of dementia was 11.56 ($SD \pm 3.75$), 45% women ($n = 27$), and 54.5% men ($n = 6$) had severe dementia. Moderate dementia was recorded in 55% women ($n = 33$) and 45.5% men ($n = 5$; interquartile range [IQR] = [8, 15]). Further sample characteristics of the nursing home residents are depicted in Table 2.

From the 46 participating registered nurses, 97.8% ($n = 45$) were female. The registered nurses had an average of 18.13 ($SD \pm 11.94$) years of professional experience. Further sample characteristics of the raters are illustrated in Table 3.

TABLE 2. Sample Characteristics of Nursing Home Residents ($N = 71$)

	n	%	Mdn	IQR	*M*	*SD*
Gender						
Women	60	84.5				
Men	11	15.5				
Age			91.00	(86, 93)	89.55	±5.85
74–84 years	14	19.7				
85–95 years	48	67.6				
96–106 years	9	12.7				
Number of residents in nursing home (NH)						
NH1	52	73.2				
NH2	12	16.9				
NH3	7	9.9				
Duration of stay			4.00	(3, 7)	5.42	±4.28
1–10 years	66	93.0				
11–21 years	4	5.6				
22–32 years	1	1.4				
MMSE			12	(8, 15)	11.56	±3.75
Severe dementia	33	46.5				
Moderate dementia	38	53.5				

Note. Mdn = median; IQR = interquartile range; MMSE = Mini-Mental State Examination.

Rater-Agreement and Inter-Rater Reliability

The results of the rater-agreement and inter-rater reliability coefficients of the BISAD items are presented in Table 4. The P_a for the four items in Area 1 (*observation before mobilization*) was 63% and above. P_a was between 70.42% and 87.32% for three of four items. The κ_w (.23–.35) and $ICC_{1.1}$ values (.234–.352) demonstrated fair concordance for all four items between the raters. For the subtotal in Area 1, moderate concordance ($\kappa_w = .40$, $ICC_{1.1} = .442$) was calculated as the highest value.

The P_a for the four items in Area 2 (*observation during mobilization*) was 53% and above. P_a was between 53.52% and 56.33% for three of four items. The item *complaints during caregiving* had a P_a of 63.38%. All four items revealed fair concordance between the raters ($\kappa_w = .26–.39$). The $ICC_{1.1}$ values of three items (.402–.446) presented moderate chance-corrected agreement; accept for the item *reaction during mobilization* ($ICC_{1.1} = .259$). For the subtotals, the P_a was 32.39% with fair ($\kappa_w = .37$) and moderate ($ICC_{1.1} = .538$) concordance. The total value of BISAD presented low P_a with 25.32% ($n = 18$) and moderate agreement ($\kappa_w = .40$, $ICC_{1.1} = .538$).

The results of the Wilcoxon test demonstrated that the subtotal of Area 2 was significantly higher than the subtotal of Area 1 ($z = -6.65$, $p = .001$).

TABLE 3. Sample Characteristics of Raters ($N = 46$)

	n	%	Mdn	IQR	*M*	*SD*
Gender						
Women	45	97.8				
Men	1	2.2				
Age			47.50	(29, 51)	42.65	±11.09
20–30 years	12	26.1				
31–41 years	5	10.9				
42–52 years	22	47.8				
53–63 years	7	15.2				
Employed at nursing home			8.50	(4,13)	9.09	±6.10
1–10 years	33	71.7				
11–21 years	10	21.7				
22–32 years	3	6.5				
Professional experience			16.00	(7, 30.25)	18.13	±11.94
2–12 years	17	37.0				
13–23 years	12	26.1				
24–34 years	13	28.3				
35–45 years	4	8.7				

Note. Mdn = median; IQR = interquartile range.

DISCUSSION

To our knowledge, this is the first study examining rater-agreement and inter-rater reliability of BISAD in residents with moderate and severe dementia in three nursing homes in Austria. As in many studies, the MMSE was used to determine the progression of cognitive impairment and severity of dementia (Likar et al., 2015; Likar et al., 2013; Lukas, Niederecker, Günther, Mayer, & Nikolaus, 2013). Loss of speech production and comprehension may occur in people with severe dementia. Then the assessment with the MMSE is not reliable (Lichtner et al., 2014). For this study, residents were excluded from participation if the MMSE was not feasible to administer.

Agreement is a verification of the extent to which both raters consistently rate the items according to their observations of the resident (Gwet, 2012). Except for the assessment of the items in Area 1, *spontaneous resting position*, *movement (or mobilization) of person (in or out of bed)*, and *interaction to others*, the observer ratings were less than the required P_a of 70%. This is an indication that pain assessment in older adults with dementia based on the item's description cannot be reliable assessed by nurses with this instrument. The lower P_a for the remaining five items and subtotals indicate the nurses' difficulty in assessing the situation. The agreement could be increased by more training, precise information, and further administration rules.

The inter-rater reliability of the item and total values were fair to moderate (Landis & Koch, 1977). Despite using a standardized assessment instrument, nurses were unable

TABLE 4. Rater-Agreement and Inter-Rater Reliability of Observation Instrument for Assessing Pain in Elderly With Dementia Items

Area 1: Observation Before Mobilization					
Items	*n*	P_a%	κ_w	ICC	95% CI
Facial expression (gaze and mimic)	45	63.38	.28	.281	[.05, .48]
Spontaneous resting position	57	80.82	.30	.243	[.01, .64]
Movement (or mobilization) of person	50	70.42	.35	.352	[.13, .54]
Interaction to others	61	87.32	.23	.234	[.00, .44]
Subtotal Area 1	37	52.11	.40	.442	[.23, .61]
Area 2: Observation During Mobilization					
Items	*n*	P_a%	κ_w	ICC	95% CI
Anxious anticipation during caregiving	40	56.33	.39	.410	[.19, .58]
Reactions during mobilization	40	56.33	.26	.259	[.02, .46]
Reactions during caregiving of painful areas	38	53.52	.39	.402	[.18, .58]
Complaints during caregiving	45	63.38	.36	.446	[.23, .61]
Subtotal Area 2	23	32.39	.37	.538	[.35, .68]
Total value	18	25.32	.40	.538	[.35, .68]

Note. P_a% = absolute agreement in %; κ_w = weighted kappa; ICC = intraclass correlation coefficient; CI = confidence interval (significant $p = .024 - .001$).

to distinguish between the pain behaviors of the affected person. It is possible for nursing home residents to show slight differences that were not depicted with BISAD. Many reasons for the low κ_w values can be mentioned, such as raters selected the individual categories with different probability, lack of consistency (Gwet, 2012; Wirtz & Caspar, 2002), nurses interpreted the items differently, or used different definitions. The inter-rater reliability of the subtotals as well as the total values with κ_w .37–.40 ($ICC_{1.1}$ = .442–.538) are also too low to derive meaningful clinical decisions. Only one other study was identified that tested inter-rater reliability. The EPCA-2 (Morello et al., 2007) had a high ICC value of .877 (95% CI [.85, .89]). However, a true comparison with those study results (Morello et al., 2007) cannot be made because the sample was selected in three university hospital centers and not in nursing homes. In addition, the French study included NVC people with different pain etiologies including dementia, cancer, stroke, and different neurological disorders. A systematic review of behavioral pain assessment tools for older adults with severe dementia (Zwakhalen, Hamers, Abu-Saad, & Berger, 2006) published results from the first version of the 11-item ECPA with high ICC values (.80) for the total value than this study (Zwakhalen et al., 2006). Results comparison is also inadmissible because the ICC value cannot be assigned to a model (Wirtz & Caspar, 2002), BISAD only contains eight items, and the item number of an instrument affects the reliability (Gwet, 2002; Mayer et al., 2004). Moreover, Fischer (2012) states that the publication by Morello et al. (2007), with detail descriptions of the EPCA-2 development, was published after his empirical investigation was completed. Hence, EPCA-2 respectively BISAD was tested in

the target population (residents with dementia) for the first time, which suggests evidence for conducting the study (Fischer, 2012). The study did not generate any inter-rater reliability values for BISAD because the nursing homes did not have the personnel resource of nursing staff available (Fischer, 2012).

The fact that pain may be intensified by movement (Gibson, 2006) was demonstrated by the difference between the subtotals of Area 1 (*before mobilization*) and Area 2 (*during mobilization*). In this study, the subtotal of Area 2 was significantly higher than of Area 1. Fischer (2012) used the paired *t* test for calculating the average difference of the subtotals. The result was also highly significant ($t = 8.121$, $p = .001$, $n = 142$).

Limitations

One limitation of this study is the small sample size and sampling procedure. The planned sample of 94 nursing home residents was not achieved because the MMSE was not feasible to administer or the resident or the resident's legally accepted representative did not give informed consent. Because of the voluntary participation of the residents, nurses, institutions, and research ethics, it was impossible to select a probability sample. A further limitation of the study is the selection bias because only three nursing homes from one state were chosen. To counteract this bias, the sample was carefully recruited according to the inclusion and exclusion criteria. The information bias was countered by conducting a detailed oral and written informed consent discussion with the nursing home resident or the resident's legally accepted representative.

Relevance to Nursing Practice and Research

BISAD does not make a reliable contribution to clinical decision making based on the inter-rater reliability on item as well as on all total values. Therefore, the implementation of BISAD is currently not recommended for the investigated setting. Because this instrument is only available in French and German, future research should focus on translating the scale into English and subsequently evaluate the psychometric properties to compare the results to existing studies.

REFERENCES

Ahlsdorf, E., & Schroeder, J. (2011). Kognitiver Status bei Demenz [Cognitive status in dementia]. In B. Reuschenbach, & C. Mahler (Eds.), *Pflegebezogene Assessmentinstrumente. Internationales Handbuch fuer Pflegeforschung und praxis* [Nursing related assessment instruments. International handbook for nursing research and practice] (pp. 349–370). Bern, Switzerland: Hans Huber.

Alzheimer Europe. (2013). *2013: The prevalence of dementia in Europe—country comparisons*. Retrieved from http://www.alzheimer-europe.org/Policy-in-Practice2/Country-comparisons/The-prevalence-of-dementia-in-Europe/(language)/eng-GB

American Geriatric Society. (2002). The management of persistent pain in older persons. *Journal of the American Geriatrics Society*, *50*(Suppl. 6), S205–S224.

Arnold, B., Brinkschmidt, T., Casser, H. R., Diezemann, A., Gralow, I., Irnich, D., . . . Söllner, W. (2014). *Multimodale Schmerztherapie fuer die Behandlung chronischer Schmerzsyndrome. Ein Konsensuspapier der Ad-hoc-Kommission Multimodale interdisziplinaere Schmerztherapie der Deutschen Schmerzgesellschaft zu den Behandlungsinhalten* [Multimodal pain therapy for the treatment of chronic pain syndromes. A consensus paper of the ad hoc commission multimodal interdisciplinary pain therapy of the German pain society on treatment content]. *Schmerz*, *28*(5), 459–472.

Benner, P. (2004). Using the Dreyfus model of skill acquisition to describe and interpret skill acquisition and clinical judgment in nursing practice and education. *Bulletin of Science, Technology & Society*, *24*(3), 188–199.

Berlin Institute for Population and Development. (2011). *Demenz-Report. Wie sich die Regionen in Deutschland, Öesterreich und der Schweiz auf die Alterung der Gesellschaft vorbereiten köennen* [Dementia report. How the regions in Germany, Austria and Switzerland can prepare for the aging society]. Retrieved from http://www.berlin-institut.org/fileadmin/user_upload/Demenz/Demenz_online.pdf

Buehler, S. (2014). Mini-Mental Status-Test (MMST). Der schnelle Standardtest [Mini-Mental Status Examination (MMSE). The fast standard examination]. *Ergopraxis*, *7*(2), 34–35.

Centmayer, R., & Lahmann, N. (2014). Schmerzen [Pain]. In T. Dassen (Ed.), *Pflegeprobleme in Deutschland. Ergebnisse von 14 Jahren Forschung in Pflegeheimen und Kliniken 2001-2014* [Nursing care problems in Germany. Results of 14 years of research in nursing homes and clinics 2001-2014] (pp. 51–56). Berlin, Germany: Charité Universitaetsmedizin Berlin, Institut fuer Gesundheits- und Pflegewissenschaft.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: Sage.

Fischer, T. (2012). *Schmerzeinschaetzung bei Menschen mit schwerer Demenz. Das Beobachtungsinstrument fuer das Schmerzassessment bei alten Menschen mit schwerer Demenz* [Pain assessment in people with severe dementia. The observation instrument for assessing pain in elderly with dementia]. Bern, Switzerland: Hans Huber.

Gibson, S. J. (2006). Older people's pain. *Pain*, *14*(3), 1–4.

Gwet, K. L. (2002). Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical Methods for Inter-Rater Reliability Assessment*, *1*, 1–5.

Gwet, K. L. (2012). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (3rd ed.). Gaithersburg, MD: Advanced Analytics.

Hadjistavropoulos, T., & Craig, K. D. (2002). A theoretical framework for understanding self-report and observational measures of pain: A communications model. *Behaviour Research and Therapy*, *40*(5), 551–570.

Hilton, A., & Skrutkowski, M. (2002). Translating instruments into other languages: Development and testing processes. *Cancer Nursing*, *25*(1), 1–7.

Hoefler, S., Bengough, T., Winkler, P., & Griebler, R. (2015). *Oesterreichischer Demenzbericht 2014* [Austrian dementia report 2014]. Retrieved from https://www.sozialministerium.at/cms/site/attachments/2/7/0/CH3434/CMS1454578572018/oesterreichischer_demenzbericht_2014.pdf

Kessler, J., Markowitsch, H. J., & Denzler, P. (2000). *Mini-Mental Status Test (MMST)* [Mini-Mental State Examination (MMSE)]. Göttingen, Germany: Beltz Test GMBH.

Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lichtner, V., Dowding, D., Esterhuizen, P., Closs, S. J., Long, A. F., Corbett, A., & Briggs, M. (2014). Pain assessment for people with dementia: A systematic review of systematic reviews of pain assessment tools. *BMC Geriatrics*, *14*, 138.

Likar, R., Pipam, W., Neuwersch, S., Köstenberger, M., Pinter, G., Gatternig, C., & Marksteiner, J. (2015). Schmerzmessung bei kognitiv beeintraechtigten Patienten mit der Doloshort-Skala [Pain measurement in cognitively impaired patients with the Doloshort scale]. *Schmerz*, *29*(4), 440–444.

Likar, R., Pipam, W., Stampfer-Lackner, W., Gatternig, K., Hammerschlag, A., Kager, I., . . . Pinter, G. (2013). Schmerzmessung bei kognitiv beeintraechtigten Patienten mit der Doloplus-2-Skala [Pain measurement in cognitively impaired patients with the Doloplus-2 scale]. In G. Pinter, R. Likar, W. Schippinger, H. Janig, O. Kada, & K. Cernic (Eds.), *Geriatrische Notfallversorgung Strategien und Konzepte* [Geriatric emergency care strategies and concepts] (pp. 361–368), Vienna, Austria: Springer Verlag.

Lukas, A., Niederecker, T., Günther, I., Mayer, B., & Nikolaus, T. (2013). Self- and proxy report for the assessment of pain in patients with and without cognitive impairment. Experiences gained in a geriatric hospital. *Zeitschrift fur Gerontologie und Geriatrie*, *46*(3), 214–221.

Mayer, H., Nonn, C., Osterbrink, J., & Evers, G. (2004). Qualitaetskriterien von Assessmentinstrumenten—Cohen's Kappa als Maß der Interrater-Reliabilitaet (Teil 1) [Quality criteria of assessment instruments—Cohen's Kappa as a measure of interrater reliability (Part 1)]. *Pflege*, *17*(1), 36–46.

Morello, R., Jean, A., Alix, M., Sellin-Peres, D., & Fermanian, J. (2007). A scale to measure pain in non-verbally communicating older patients: The ECPA-2 Study of its psychometric properties. *Pain*, *133*(1), 87–98.

Nestler, N., Gnass, I., & Schuler, M. (2015). Schmerz in Einrichtungen der stationaeren Altenhilfe [Pain in facilities of inpatient elderly care]. *Schmerz*, *29*(1), 131–141.

Osterbrink, J., Hufnagel, M., Kutschar, P., Mitterlehner, B., Krüger, C., Bauer, Z., . . . Ewers, A. (2012). Die Schmerzsituation von Bewohnern in der stationaeren Altenhilfe. Ergebnisse einer Studie in Muenster [The painful situation of residents in the inpatient elderly. Results of a study in Muenster]. *Schmerz*, *26*(1), 27–35.

Schreier, M., Stering, U., Pitzer, S., Iglseder, B., & Osterbrink, J. (2015). Schmerz und Schmerzerfassung in Altenpflegeheimen. Ergebnisse der OSiA-Studie [Pain and pain assessment in elderly nursing care homes. Results of the OSiA-Study]. *Schmerz*, *29*(2), 203–210.

Snow, A. L., O'Malley, K. J., Cody, M., Kunik, M. E., Ashton, C. M., Beck, C., . . . Novy, D. (2004). A conceptual model of pain assessment for noncommunicative persons with dementia. *The Gerontologist*, *44*(6), 807–817.

Snow, A. L., & Shuster, J. L., Jr. (2006). Assessment and treatment of persistent pain in persons with cognitive and communicative impairment. *Journal of Clinical Psychology*, *62*(11), 1379–1387.

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, *9*(4), 1–11. Retrieved from http://PAREonline.net/getvn.asp?v=9&n=4

Thomm, M. (Ed.). (2011). *Schmerzmanagement in der Pflege* [Pain management in nursing care]. Berlin, Germany: Springer Verlag.

Wirtz, M., & Caspar, F. (2002). *Beurteileruebereinstimmung und Beurteilerreliabilitaet* [Rater-agreement and rater-reliability]. Bern, Switzerland: Hogrefe.

World Health Organization. (2012). *Dementia: A public health priority*. Retrieved from http://www.who.int/mental_health/publications/dementia_report_2012/en/

Zwakhalen, S., Hamers, J., Abu-Saad, H. H., & Berger, M. (2006). Pain in elderly people with severe dementia: A systematic review of behavioural pain assessment tools. *BMC Geriatrics*, *6*, 3.

Acknowledgments. We thank the participating nursing homes, patients, and especially the registered nurses who supported us during data collection.

Correspondence regarding this article should be directed to Gerhard Mueller, Assoc.-Prof. Dr., RN, UMIT-Private University for Health Sciences, Medical Informatics and Technology, Department of Nursing Science and Gerontology, Eduard Wallnoefer-Zentrum 1, A-6060 Hall in Tyrol, Austria. E-mail: gerhard.mueller@umit.at