EUROPEAN GERIATRIC MEDICINE

Title: Psychometric properties of pain measurements for people living with dementia: a COSMIN systematic review.

Authors: Toby O Smith,[1,2] Karmen Harvey[2]

Affiliation

1. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, OX3 7LD, UK
2. Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Corresponding Author: Dr Toby Smith, Botnar Research Centre, NDORMS, University of Oxford, OX3 7LD, UK. Email: toby.smith@ndorms.ox.ac.uk

DECLARATIONS AND ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS: None

COMPETING INTERESTS OR FUNDING: No author disclose any financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

FUNDING: No funding was received to conduct this work.

ETHICS APPROVAL: Ethical approval was not required in relation to this systematic review.

AUTHORSHIP: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Toby Smith and Karmen Harvey. The first draft of the manuscript was written by Toby Smith and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

KEY SUMMARY POINTS

AIM: To determine the psychometric properties of the most frequently used pain measurement tools in research of people living with dementia.

FINDINGS: There was strong and moderate level evidence to support the use of the facial action coding system, PACSLAC and PACSLAC-II, CNPI, DOLOPLUS-2, ALGOPLUS, MOBID and MOBID-2 tools for the assessment of pain with people living with dementia. There was limited evidence to support the use of the Abbey Pain Scale, PAINAD and self-reported pain through verbal rating pain score.

MESSAGE: This study has identified which outcome measures are the most robust to assess pain in older people with dementia.

ABSTRACT

PURPOSE: Detecting pain in older people with dementia is challenging. Consequentially, pain is often under-reported and under-treated. There remains uncertainty over what measures should be promoted for use to assess pain in this population. The purpose of this paper is to answer this question.

METHODS: A search of clinical trials registered on the ClinicalTrial.gov and ISRCTN registries was performed to identify outcome measures used to assess pain in people with dementia. Following this, a systematic review of published and unpublished databases was performed to 01 November 2021 to identify papers assessing the psychometric properties of these identified measures. Each paper and measure was assessed against the COSMIN checklist. A best evidence synthesis analysis was performed to assess the level of evidence for each measure.

RESULTS: From 188 clinical trials, nine outcome measures were identified. These included: Abbey Pain Scale, ALGOPLUS, DOLOPLUS-2, Facial Action Coding System, MOBID-2, self-reported pain through the NRS or VAS/thermometer or Philadelphia Geriatric Pain Intensity Scale, PACSLAC/PACSLAC-2, Pain Assessment in Advanced Dementia (PAINAD), Checklist for non-verbal pain behavior (CNPI). From these 51 papers (5924 people with dementia) were identified assessing the psychometric properties ot these measures. From these, there was strong and moderate level evidence to support the use of the facial action coding system, PACSLAC and PACSLAC-II, CNPI, DOLOPLUS-2, ALGOPLUS, MOBID and MOBID-2 tools for the assessment of pain with people living with dementia.

CONCLUSION: Whilst these reflect measurement tools used in research, further consideration on how these reflect clinical practice, should be considered.

Keywords: Pain; Distress; Outcome Measure; Instrument; Older People; Cognitive Impairment

PROSPERO Registration: CRD42021282032

INTRODUCTION

Dementia is a major, worldwide health challenge predominantly affecting older people. It has an estimate global prevalence of 45 million people [1]. Pain is frequently reported in older people with approximately 20% to 50% living with chronic pain [2]. Managing pain can be difficult. There are challenges surrounding adherence and adoption of interventions such as exercise and medication taking. Detecting pain can also be difficult for people with dementia. Accordingly, pain in people with dementia is often under-detected and under-treated [3].

Self-reported pain scales such as numerical rating scales (NRS) are most frequently used to assess pain. For these patients, self-reported pain alone may not be sufficient [3]. Observed behavioural indicators of pain such as verbal complaints, sighing, moaning, agitation, crying, grimacing, rapid blinking, restlessness, rubbing, disorientation or aggression may be valuable [4,5].

Lichtner et al [6] previously identified eight literature reviews reporting measurements and psychometric properties of tools assessing pain in people with dementia. No single tool was identified as more reliable and valid than others, with a wide variation in the reliability and validity. However the search from the most recent review was performed in 2013. Furthermore no studies have assessed the psychometric properties of outcome measures against the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. This is a major limitation as the COSMIN checklist[7] is a robust assessment of both methodological quality of studies assessing measurement properties, with the quality of the outcome measure itself. Through this, the COSMIN checklist offers a robust, evidence-based recommendation on the quality of outcome measures selection in research and clinical practice [7].

The assessment of pain using a valid and accurate measurement is the basis for successful pain management [8]. However there remains uncertainty on the appropriateness of these measures. Accordingly, the purpose of this systematic review was to determine the psychometric properties of the most frequently used pain measurement tools in research of people living with dementia.

METHODS

This systematic review was conducted according to the COSMIN guidance[7] and reported in accordance with the PRISMA statement [9]. The study protocol was registered prior to commencing (PROSPERO registration: CRD42021282032).

Search Strategy

Search 1: To identify the measurement tools currently used to measure pain in clinical trials of people living with dementia, we performed a search of the databases ClinicalTrial.gov and ISRCTN from inception to 01 October 2021. We used the search terms “Dementia OR cognitive impairment” AND “pain”.

Search 2: A systematic review was undertaken of published and unpublished sources to identify potentially eligible studies assessing the psychometric properties of pain measurement tools identified from Search 1. We searched the published databases: Medline, CINHAL, EMBASE, AMED, PsycINFO and DARE from database inception to 01 November 2021. We also searched the trial

registry and unpublished literature databases OpenGrey, Clinicaltrials.gov and ISRCTN registries from inception to 01 November 2021. The search terms used for the EMBASE database are presented in **Supplementary File 1**. These were based on the COSMIN search filters to identify studies of psychometric properties linked to terms related to dementia, cognitive impairment and pain. The search strategy was optimised for each electronic database search. The reference lists of all potentially eligibility studies were reviewed and the corresponding authors from each included study were contacted and asked to review the search results.

Eligibility Assessment

For both Search 1 and 2, studies were included if they recruited people, aged 60 years and older, with dementia. Dementia criteria such as the Diagnostic and Statistical Manual of Mental Disorders, Revised Fourth Edition (DSM IV) [10], National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) [11], the National Institute for Neurological Disorders and Stroke-Association Internationale pour la Recherche et Enseignement en Neurosciences (NINCDS-AIREN)[12] were considered appropriate. Where self-reported dementia was reported, further scrutiny of the characteristics of the population in relation to severity of cognitive impairment, age and comorbidities were considered. Where uncertain, corresponding authors were asked to verify the approach used to define dementia. All stages and severities of dementia were eligibility i.e. mild, moderate, severe. Whilst it is acknowledged that pain assessment tools have been developed for other, non-dementia, patient groups with cognitive impairment [13], these were excluded from this review unless there was sufficient evidence that participants presented with dementia.

We did not restrict the form, cause or pathology causing pain. Through this, participant's pain arise from musculoskeletal, post-surgery, medical and cancer-related sources.

We included studies regardless of setting i.e. acute, community, residential or nursing home. We excluded studies not published in English, narrative and systematic reviews, although reviewed the reference lists of these publications to identify any previously omitted studies.

For Search 2, we included all full-text publications which reported any assessment of the psychometric properties of measurement tools identified from Search 1. Papers which included findings on pain management were considered if they also provided data on the psychometric properties of a pain measurement tool. We only included studies which reported one or more of the COSMIN taxonomy of: internal consistency, test-retest reliability, measurement error, content validity, structural validity, construct validity/hypotheses testing, cross-cultural validity, criterion validity or responsiveness [7].

Study Identification

The search results were screened against the eligibility criteria by two reviewers (TS, KH). This was initially by title and abstract, and then by full-text version. Screening was performed by each reviewer independently. When consensus on study eligibility could not be reached, agreement was reached through discussion.

Data Extraction

For each included study, data were extracted independently by one reviewer (TS). This was then verified for accuracy by a second reviewer (KH). Where disagreements occurred, these were resolved through discussion.

Data were extracted onto a bespoke data extraction table. Data extracted included: measurement tool name, setting tested, country of assessment, method of administration, person administered, duration between testing (if appropriate), patient participant characteristics (number and response rate), age, gender, diagnosis of pain, diagnosis of dementia, severity of dementia), psychometric outcomes (reliability, validity, responsiveness).

Risk of Bias

To assess the methodological quality of the included studies, the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist[14] was used. The COSMIN checklist assesses the following measurement properties: content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity and responsiveness. The overall quality of how each measurement property was evaluated on a four-point scale: very good, adequate, doubtful or inadequate, as per the COSMIN guidance. The methodological quality score per property was then obtained by taking the lowest rating of any item in each box – worst score counts principle. Two reviewers (TS,KH) assessed each study using this approach independently with disagreements resolved through consensus.

Data Analysis

The psychometric properties of each measurement tool were reported narratively. Through this descriptive statistics, inferential statistics and degrees of variance were reported from included studies. Analysis was made following Chiarotto et al[15] best evidence synthesis approach where 'strong' was a measurement tool which demonstrate consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality; 'moderate' demonstrated consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality, 'limited' demonstrated on study of fair methodological quality, conflicting demonstrated conflicting findings and 'unknown' was only for studies of poor methodological quality or no studies reporting a measure.

RESULTS

Search 1: Identification of measurement tools

In total, 188 individual clinical trials were identified from Search 1. Of these, 56 were identified which reported measuring pain with participants living with dementia. A summary of these studies is presented in **Table 1**.

From the list generated from Search 1, we excluded all measures which did not specifically assess pain but included pain as a sub-domain of an instrument e.g. SF-36, WOMAC and EQ-5D. From this, seven outcomes were excluded (Comfort Assessment in Dying with Dementia, Edmonton Symptom Assessment Scale, EQ-5D, GLOBAL PROMIS-10, SF-36, Resident Assessment Index-Minimum Dataset, Symptom Management - End of Life for Dementia). We excluded measurement tools which were not designed for people with cognitive impairment. Accordingly three instruments were excluded (Brief Pain Inventory, McGill Pain Map, WOMAC). Resultantly, the psychometric properties of nine measurement tools formed the basis of Search 2 (Abbey Pain Scale, ALGOPLUS, DOLOPLUS-2, Facial Action Coding System, MOBID-2, self-reported pain through the NRS or VAS/thermometer or

Philadelphia Geriatric Pain Intensity Scale, PACSLAC/PACSLAC-2, Pain Assessment in Advanced Dementia (PAINAD), Checklist for non-verbal pain behavior (CNPI)(**Supplementary File 2**).

Search 2: Psychometric Tools Analysis

A summary of the Search 2 results is presented in **Figure 1**. In total, 1173 individual citations were identified. Fifty-one studies reported data on the psychometric properties of one or more of the nine measurement tools identified in Search 1. These studies were included in the analysis.

Characteristics of included studies and quality assessment

A summary of the characteristics of the included studies is presented in **Table 2**. In total, 5924 people with dementia were assessed. Mean age of population ranged from 72.5 years[16] to 87.9 years [17]. Thirteen studies were performed in a hospital setting [16, 18-29], 33 in care home facilities [17, 30-61] and two studies were based in both care home and people's home settings [62, 63]. Two studies were performed both in care home and hospital settings [64,65]. The location of study was not stated in Lorenzet et al [66]. Studies were reported in 21 countries, most frequently Norway (n=8) [32,41,48,56-59,63], USA (n=7) [19,33,34,42,44,60,61], Canada (n=4) [31,52,54,55] and Brazil (n=4)]17,22,23,66].

A summary of the findings from the COSMIN assessment is presented in **Supplementary File 3**. The results for the psychometric analysis are presented in **Supplementary File 4**. A summary of findings for the best evidence synthesis is presented as **Table 3**.

Abbey Pain Scale

Eight studies reported data on the psychometric properties of the Abbey Pain Scale [35-40,43,46]. Overall there was limited evidence for the use of the Abbey Pain Scale (**Table 3**). There was inadequate evidence on PROM development, internal consistency (Cronbach: 0.65-0.74), cross-cultural validity and responsiveness ($p<0.001$). There was adequate evidence for the assessment of construct validity (R=0.49-0.91) and very good evidence for reliability (inter-rater: 0.75-0.88; intra-rater: 0.66-0.68). The level of evidence for structural validity was doubtful (Cronbach: 0.76).

Pain Assessment in Advanced Dementia (PAINAD)

Twelve studies assessed the PAINAD [16,20-24,40,44-47,65]. Overall the level of evidence for the PAINAD tool was limited (**Table 3**). Whilst there was an adequate level of evidence for construct validity (R=0.48-0.88), very good level of evidence for internal consistency (Cronbach alpha: 0.65-0.84) and reliability (intra-rater: 0.71-0.89; inter-rater: 0.79-0.94), there was inadequate evidence for cross-cultural validity and responsiveness ($p<0.001$). There was doubtful level of evidence for structural validity (variance explained: 46.5% to 68.9%).

Facial Action Coding System

Five studies provided data on the facial action coding system [18,27,30,31,64]. These demonstrated moderate evidence for the use of this measurement tool (**Table 3**). There was adequate evidence for construct validity (R=0.116-0.463), structural validity ($p=0.06$ to $p<0.001$) and reliability (inter-rater: 0.94)

Checklist for non-verbal pain behavior (CNPI)

Six studies presented data on the psychometric properties of the CNPI [19,41-44,55]. Overall there was moderate evidence for the CNPI (**Table 3**). There was adequate evidence for construct validity (R=0.46-0.88) and very good evidence of reliability (intra-rater: 0.23-0.65; inter-rater: 0.45-0.59). However there was inadequate evidence for internal consistency (Cronbach alpha: 0.64-0.90).

Self-reported pain through verbal rating pain score

Ten studies assessed the psychometric properties of self-reported/verbal rating pain measures [27-29,33-35,42,45,51,54]. Overall there was limited evidence supporting the use of these tools (**Table 3**). Whilst there was adequate evidence on PROM development, construct validity (R=0.30 to 0.95) and reliability (intra-rater: 0.71-0.84; inter-rater: 0.81-0.97), there was inadequate evidence on internal consistency (Cronbach: 0.74-0.84) and responsiveness (p=0.03).

ALGOPLUS

One study, performed in a French hospital setting, presented data on the psychometric properties of the ALGOPLUS instrument [29]. This provided strong evidence for this tool (**Table 3**). Data reported very high construct validity (r2=0.81; p<0.001), very high inter-rater reliability (0.812) and internal validity (KR-20: 0.712) and responsiveness to treatment (p<0.001).

MOBID and MOBID-2

Four studies presented data on the psychometric properties of the MOBID [56,58,60,61]. Overall, the MOBID instruments demonstrated moderate evidence (**Table 3**). If offered adequate evidence for PROM development and construct validity (R=0.51-0.54 [60,61]. Whilst the instrument demonstrated doubtful evidence for internal consistency the values were high (Cronbach: 0.83-0.89), it demonstrated adequate evidence for reliability (inter-rater: 0.86-0.97; intra-rater: 0.79-0.92).

Two studies reported data on the MOBID-2 [57,59] instrument. It demonstrated moderate evidence for use (**Table 3**). There was adequate evidence for PROM development and construct validity (R=0.61), measurement error (Standard Error of Measurement (SEM): 1.4). Whilst there was inadequate evidence for the responsiveness, the minimally clinically important difference (MCID) was reported as three points and reported to be responsive to treatment (p<0.001). There was very good evidence for the MOBID-2 for internal consistency (Cronbach: 0.82-0.84) and reliability (inter-rater: 0.94; intra-rater: 0.85-0.92).

PACSLAC and PACSLAC-II

Four studies assessed the PACSLAC-II [30,31,55,62]. They suggested moderate evidence to support the use of this measurement tool (**Table 3**). There was very good evidence for internal consistency (Cronbach: 0.74-0.77), and reliability (inter-rater: 0.63-0.86) and adequate evidence for construct validity (R=0.54-0.68). However, there was inadequate evidence for the assessment of responsiveness (p<0.01).

The PACLAC was assessed in six studies [17,40,52,53,54,66]. This demonstrated moderate evidence (**Table 3**). There was very good evidence for PROM development. There was adequate evidence for construct validity (R=0.54-0.72), internal consistency (Cronbach alpha: 0.77-0.87), reliability (inter-rater: 0.52-0.96; intra-rater: 0.86) and responsiveness (p<0.001). There was doubtful evidence for structural validity and cross-cultural validity.

DOLOPLUS-2

Thirteen studies assessed the psychometric properties of the DOLOPLUS-2 [25-28,32,44,46,48-51,62,63]. Overall there was moderate evidence to support the use of this measurement tool. It demonstrated very good evidence for the assessment of internal consistency (Cronbach: 0.770 to 0.95) and reliability (intra-rater: 0.71; inter-rater: 0.35-0.86). There was adequate evidence for construct validity (R=0.33-0.70), measurement error (SEM: ±1.759) and cross-cultural validity. There was doubtful evidence for structural validity (explained variance: 36.9% to 76.1%) and inadequate evidence on responsiveness ($p<0.001$).

DISCUSSION

The findings indicate strong and moderate evidence to support the use of the facial action coding system, PACSLAC and PACSLAC-II, CNPI, DOLOPLUS-2, ALGOPLUS, MOBID and MOBID-2 tools. There is limited evidence for the Abbey Pain Scale, self-reported pain measures and the PAINAD tool.

The literature highlights the challenges of assessing pain with people living with dementia [3,4,67]. Challenges have included insufficient time to use measurement tools [68,69], user's uncertainty over the reliability of these [70], access to physically finding and using the measurement tools [71], and perceived superiority of observational methods of behaviours and physical manifestations of pain [70]. Whilst there is a bias to observational manifestation in a number of the supported measurement tools recommended, the time to complete and interpret these may act as a further barrier to adoption. Consideration of such potential challenges may be made when exploring the implementation of recommended measurement tools.

Under-treatment of pain in people with dementia has been attributed to challenges in recognition and assessment of pain, coupled with reservations on polypharmacy and side effects of analgesia [72]. Achterberg et al [73] highlighted the frequently seen scenario where people with dementia are prescribed analgesics, but due to concerns around side effects, particularly regarding non-steroidal anti-inflammatory drugs, opioids and adjunct analgesics, the medications are either not administer or are at a sufficient dosage to manage symptoms. This was clearly illustrated in Roitto et al's[74] survey where although 19% of their 327 cohort of people living in nursing homes with dementia were prescribed opioids, 79% were still in pain. Whilst this study has highlighted potentially robust pain measurement tools for this population, implementing both the assessment and subsequent treatment to improve pain management is required.

Pain assessment ideally considers several pain dimensions. These include: intensity, location, affect, cognition, behaviour and social accompaniments [72]. Measurement tools, most notably the DOLOPLUS-2, are multi-dimensional. Conversely, self-reported VAS/NRS of observation are unidimensional. However, it is acknowledged that assessment of some dimensions, notably pain cognition, can be more challenging due to communication and cognitive barriers. Focusing on single dimensions should be avoided to negate the risks of under-reporting/under-representing pain experienced by individuals.

Whilst reliability and construct validity were well-explored, there remains limited evidence of the responsiveness, structural validity and measurement error for many of the identified measures. This may be a reason for why pain measurement tools are poorly adopted into practice. Improving confidence around how measurement tools are used and interpreted may promote the implementation of such tools. Furthermore, as observational tools were most widely assessed,

understanding the 'normal' or familiar behaviours of a person with dementia is important to recognise when something abnormal or noxious is being felt. No studies assessed the difference in reliability or validity when the assessment was performed by a healthcare professional versus a close relative or friend who may be more familiar with the individual. This may be an important area for future study, particularly when considering the adoption of pain assessment instruments in community and non-health or social care profession settings.

This systematic review presents with a number of strengths and limitations. A major strength is the adoption of the COSMIN evaluation. This approach ensured the reader could be fully informed on the confidence with the recommendations made based on the evidence. Three important limitations should be considered. Firstly, a comprehensive approach to reporting the psychometric properties of the most frequently used measurement instruments in research was adopted to aid prioritisation. However, this meant measurement tools used in clinical practice but not trials, may have been omitted. Secondly given the methods adopted through Search 1 to identify potential measurement tools, more recent tools such as the ePAT were not included in the analysis [39]. Consideration of this and inclusion of forthcoming evidence on psychometric properties should be made to update the findings as new evidence evolves in the field. Secondly, there was insufficient evidence to assess differences in recommendations based on severity of dementia. Evaluation on the impact of severity of cognitive impairment on the performance of the identified measurement tools would be warranted. Finally, there were challenges cause by poor reporting within included studies. There was insufficient detail within included studies to ascertain whether pain assessment instruments assessed acute or chronic pain, or whether individuals were taking analgesia or not. This may impact on the generalisability of the findings into practice and should be consider when reporting future studies in this area.

To conclude, there is strong and moderate evidence to support the use of the facial action coding system, PACSLAC and PACSLAC-II, CNPI, DOLOPLUS-2, ALGOPLUS, MOBID and MOBID-2 tools for the assessment of pain with people living with dementia. Whilst these reflect measurement tools used in research, further consideration on how these reflect clinical practice, and lessons on how to implement these tools into practice should be considered to improve the detection and management of pain for people with dementia.

REFERENCES

1. Prince M, Ali GC, Guerchet M, Prina AM, Albanese E, Wu YT. Recent global trends in the prevalence and incidence of dementia, and survival with dementia. Alzheimers Res Ther. 2016;8:23

2. Apinis C, Tousignant M, Arcand M, Tousignant-Laflamme Y. Can adding a standardized observational tool to interdisciplinary evaluation enhance the detection of pain in older adults with cognitive impairments? Pain Med. 2014;15:32-41.

3. Zwakhalen SM, Hamers JP, Abu-Saad HH, Berger MP. Pain in elderly people with severe dementia: a systematic review of behavioural pain assessment tools. BMC Geriatr. 2006;6:3.

4. Malara A, De Biase GA, Bettarini F, Ceravolo F, Di Cello S, Garo M, Praino F, Settembrini V, Sgrò G, Spadea F, Rispoli V. Pain Assessment in Elderly with Behavioral and Psychological Symptoms of Dementia. J Alzheimers Dis. 2016;50:1217-25.

5. American Geriatrics Society. AGS panel on the pharmacological management of persistent pain in older persons. J Am Geriatr Soc 2009;57:1331–4.

6. Lichtner V, Dowding D, Esterhuizen P, Closs SJ, Long AF, Corbett A, Briggs M. Pain assessment for people with dementia: a systematic review of systematic reviews of pain assessment tools. BMC Geriatr. 2014;14:138.

7. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27:1147-57.

8. Corbett A, Husebo B, Malcangio M, Staniland A, Cohen-Mansfield J, Aarsland D, Ballard C. Assessment and treatment of pain in people with dementia. Nat Rev Neurol. 2012;8:264-74.

9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71

10. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fourth edition, text revision (DSM-IV-TR). American Psychiatric Association, Washington, DC; 2000.

11. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of department of health and human services task force on Alzheimer's Disease. Neurology. 1984;34:934–94.

12. Roman GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, Amaducci L, Orgogozo JM, Brun A, Hofman A et al. Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. Neurology. 1993;43:250–60.

13. Decker SA, Perry AG. The development and testing of the PATCOA to assess pain in confused older adults. Pain Manag Nurs. 2003;4:77-86.

14. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10:22.

15. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. Phys Ther. 2016;96:1620-37.

16. Büyükturan O, lkin Naharci M, Büyükturan B, Kirdi N, Yetiş A. The Turkish Version of Pain Assessment in Advanced Dementia (PAINAD) Scale. Noro Psikiyatr Ars. 2018;55:271–75.

17. Thé KB, Gazoni FM, Cherpak GL, Lorenzet IC, Alves dos Santos L, Nardes EM, dos Santos FC. Pain assessment in elderly with dementia: Brazilian validation of the PACSLAC scale. Einstein 2016; 14:152–7.

18. Kunz M, Scharmann S, Hemmeter U, Schepelmann K, Lautenbacher S. The facial expression of pain in patients with dementia. Pain. 2007;133,221–8.

19. Feldt KS. The checklist of nonverbal pain indicators (CNPI). Pain Manag Nurs. 2000;1:13-21.

20. Costardi D, Rozzini L, Costanzi C, Ghianda D, Franzoni S, Padovani A, Trabucchi M. The Italian version of the pain assessment in advanced dementia (PAINAD) scale. Arch Gerontol Geriatr. 2007;44:175-80.

21. Batalha LMC; Duarte CIA, do Rosário RAF, da Costa MFSP, Pereira VJR, Morgado TMM. Adaptação cultural e propriedades psicométricas da versão portuguesa da escala Pain Assessment in Advanced Dementia. Rev Esc Enferm USP. 2012;8:7-16.

22. Valera GG, Carezzato NL, Vale FA, Hortense P. Cultural adaptation of the scale Pain Assessment in Advanced Dementia - PAINAD to Brazil. Rev Esc Enferm USP. 2014;48:462-8.

23. Pinto MC, Minson FP, Lopes AC, Laselva CR. Cultural adaptation and reproducibility validation of the Brazilian Portuguese version of the Pain Assessment in Advanced Dementia (PAINAD-Brazil) scale in non-verbal adult patients. Einstein. 2015;13:14-9.

24. Cantón-Habas V, Carrera-González MDP, Moreno-Casbas MT, Rich-Ruiz M. Spanish adaptation and validation of the Pain Assessment Scale in Advanced Dementia (PAINAD) in patients with dementia and impaired verbal communication: cross-sectional study. BMJ Open. 2021;11:e049211.

25. Ando C, Hishinuma M. Development of the Japanese DOLOPLUS-2: a pain assessment scale for the elderly with Alzheimer's disease. Psychogeriatrics. 2010;10:131-7.

26. Ando C, Ito Y, Amemiya S, Tamura K, Kako K, Tsuzura S, Yoshida R, Hishinuma M. Effectiveness of the Japanese DOLOPLUS-2: a pain assessment scale for patients with moderate-to-severe dementia. Psychogeriatrics. 2016;16:315-22.

27. Pautex S, Herrmann F, Le Lous P, Fabjan M, Michel JP, Gold G. Feasibility and reliability of four pain self-assessment scales and correlation with an observational rating scale in hospitalized elderly demented patients. J Gerontol A Biol Sci Med Sci. 2005;60:524-9.

28. Pautex S, Herrmann FR, Michon A, Giannakopoulos P, Gold G. Psychometric properties of the Doloplus-2 observational pain assessment scale and comparison to self-assessment in hospitalized elderly. Clin J Pain. 2007;23:774-9.

29. Rat P, Jouve E, Pickering G, Donnarel L, Nguyen L, Michel M, Capriz-Ribière F, Lefebvre-Chapiro S, Gauquelin F, Bonin-Guillaume S. Validation of an acute pain-behavior scale for older persons with inability to communicate verbally: Algoplus. Eur J Pain. 2011;15:198.e1-198.e10.

30. Erin Browne M, Hadjistavropoulos T, Prkachin K, Ashraf A, Taati B. Pain expressions in dementia: validity of observers' pain judgments as a function of angle of observation. J Nonverbal Behav. 2019;43:309-27.

31. Hadjistavropoulos T, Browne ME, Prkachin KM, Taati B, Ashraf A, Mihailidis A. Pain in severe dementia: A comparison of a fine-grained assessment approach to an observational checklist designed for clinical settings. Eur J Pain. 2018;22:915-25.

32. Hølen JC, Saltvedt I, Fayers PM, Bjørnnes M, Stenseth G, Hval B, Filbet M, Loge JH, Kaasa S. 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.

33. Parmelee PA, Smith B, Katz IR. Pain complaints and cognitive status among elderly institution residents. J Am Geriatr Soc. 1993;41:517-22.

34. Weiner D, Peterson B, Keefe F. Evaluating persistent pain in long term care residents: what role for pain maps? Pain. 1998;76:249-57.

35. Abbey J, Piller N, De Bellis A, Esterman A, Parker D, Giles L, Lowcay B. The Abbey pain scale: a 1-minute numerical indicator for people with end-stage dementia. Int J Palliat Nurs. 2004;10:6-13.

36. Takai Y, Yamamoto-Mitani N, Chiba Y, Nishikawa Y, Hayashi K, Sugai Y. Abbey Pain Scale: development and validation of the Japanese version. Geriatr Gerontol Int. 2010;10:145-53.

37. van Iersel T, Timmerman D, Mullie A. Introduction of a pain scale for palliative care patients with cognitive impairment. Int J Palliat Nurs. 2006;12:54-9.

38. Babicova I, Cross A, Forman D, Hughes J, Hoti K. Evaluation of the Psychometric Properties of PainChek® in UK Aged Care Residents with advanced dementia. BMC Geriatr. 2021;21:337.

39. Atee M, Hoti K, Parsons R, Hughes JD. Pain Assessment in Dementia: Evaluation of a Point-of-Care Technological Solution. J Alzheimers Dis. 2017;60:137-50.

40. Liu JY, Briggs M, Closs SJ. The psychometric qualities of four observational pain tools (OPTs) for the assessment of pain in elderly people with osteoarthritic pain. J Pain Symptom Manage. 2010;40:582-98.

41. Nygaard HA, Jarland M. The Checklist of Nonverbal Pain Indicators (CNPI): testing of reliability and validity in Norwegian nursing homes. Age Ageing. 2006;35:79-81.

42. Ersek M, Polissar N, Neradilek MB. Development of a composite pain measure for persons with advanced dementia: exploratory analyses in self-reporting nursing home residents. J Pain Symptom Manage. 2011;41:566-79.

43. Neville C, Ostini R. A psychometric evaluation of three pain rating scales for people with moderate to severe dementia. Pain Manag Nurs. 2014;15:798-806.

44. Ersek M, Herr K, Neradilek MB, Buck HG, Black B. Comparing the psychometric properties of the Checklist of Nonverbal Pain Behaviors (CNPI) and the Pain Assessment in Advanced Dementia (PAIN-AD) instruments. Pain Med. 2010;11:395-404.

45. 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;35:252-6.

46. Zwakhalen SM, Hamers JP, Berger MP. The psychometric quality and clinical usefulness of three pain assessment tools for elderly people with dementia. Pain. 2006;126:210-20.

47. Lin PC, Lin LC, Shyu YI, Hua MS. Chinese version of the Pain Assessment in Advanced Dementia Scale: initial psychometric evaluation. J Adv Nurs. 2010;66:2360-8.

48. Torvik K, Kaasa S, Kirkevold O, Rustøen T. Pain and quality of life among residents of Norwegian nursing homes. Pain Manag Nurs. 2010;11:35-44.

49. Chen YH, Lin LC, Watson R. Validating nurses' and nursing assistants' report of assessing pain in older people with dementia. J Clin Nurs. 2010;19:42-52.

50. Akbarzadeh M, Jakobsson U. Assessing pain mong older people with communication difficulties – a psychometric evaluation of DOLOPLUS-2. Vard I Norden. 2007;86:26–31.

51. Chen YH, Lin LC, Watson R. Evaluation of the psychometric properties and the clinical feasibility of a Chinese version of the Doloplus-2 scale among cognitively impaired older people with communication difficulty. Int J Nurs Stud.2010;47:78-88.

52. Fuchs-Lacelle S, Hadjistavropoulos T. Development and preliminary validation of the pain assessment checklist for seniors with limited ability to communicate (PACSLAC). Pain Manag Nurs. 2004;5:37-49.

53. Cheung G, Choi P. The use of the Pain Assessment Checklist for Seniors with Limited Ability to Communicate (PACSLAC) by caregivers in dementia care. N Z Med J. 2008;121:21-9.

54. Kaasalainen S, Akhtar-Danesh N, Hadjistavropoulos T, Zwakhalen S, Verreault R. A comparison between behavioral and verbal report pain assessment tools for use with residents in long term care. Pain Manag Nurs. 2013;14:e106-e114.

55. Chan S, Hadjistavropoulos T, Williams J, Lints-Martindale A. Evidence-based development and initial validation of the pain assessment checklist for seniors with limited ability to communicate-II (PACSLAC-II). Clin J Pain. 2014;30:816-24.

56. Husebo BS, Strand LI, Moe-Nilssen R, Husebo SB, Snow AL, Ljunggren AE. Mobilization-Observation-Behavior-Intensity-Dementia Pain Scale (MOBID): development and validation of a nurse-administered pain assessment tool for use in dementia. J Pain Symptom Manage. 2007;34:67-80.

57. Husebo BS, Strand LI, Moe-Nilssen R, Husebo SB, Ljunggren AE. Pain in older persons with severe dementia. Psychometric properties of the Mobilization-Observation-Behaviour-Intensity-Dementia (MOBID-2) Pain Scale in a clinical setting. Scand J Caring Sci. 2010;24:380-91.

58. Husebo BS, Strand LI, Moe-Nilssen R, Husebo SB, Ljunggren AE. Pain behaviour and pain intensity in older persons with severe dementia: reliability of the MOBID Pain Scale by video uptake. Scand J Caring Sci. 2009;23:180-9.

59. Husebo BS, Ostelo R, Strand LI. The MOBID-2 pain scale: reliability and responsiveness to pain in patients with dementia. Eur J Pain. 2014;18:1419-30.

60. Herr K, Sefcik JS, Neradilek MB, Hilgeman MM, Nash P, Ersek M. Psychometric Evaluation of the MOBID Dementia Pain Scale in U.S. Nursing Homes. Pain Manag Nurs. 2019;20:253-60.

61. Sefcik J, Herr K, Neradilek M, Hilgeman M, Nash P, Ersek M. Psychometrics of the mobility observation-behaviour-intensity-dementia (MOBID) pain scale in US nursing homes. Innovation in Aging 2018;S1:635-6.

62. Zare M, Tagharrobi Z, Sharifi K, Sooki Z, Abolhasani J. Psychometric evaluation of the Persian version of the Doloplus-2 (P-Doloplus-2) scale in elderly with dementia. Turk J Med Sci. 2020;50:953-62.

63. Hølen JC, Saltvedt I, Fayers PM, Hjermstad MJ, Loge JH, Kaasa S. Doloplus-2, a valid tool for behavioural pain assessment? BMC Geriatr. 2007;7:29.

64. Lautenbacher S, Walz AL, Kunz M. Using observational facial descriptors to infer pain in persons with and without dementia. BMC Geriatr. 2018;18:88.

65. Cantón-Habas V, Rich-Ruiz M, Moreno-Casbas MT, Ramírez-Expósito MJ, Martínez-Martos JM, Carrera-González MDP. Correlation between Biomarkers of Pain in Saliva and PAINAD Scale in Elderly People with Cognitive Impairment and Inability to Communicate. J Clin Med. 2021;10:1424.

66. Lorenzet IC, Dos Santos FC, De Souza PMR, Gambarro RC. Assessment of pain in elderly patients with dementia: translation and transcultural adaptation of the instrument PACSLAC into Portuguese. Revista Brasileira de Medicina. 2011;68:129-34.

67. Bullock L, Chew-Graham CA, Bedson J, Bartlam B, Campbell P. The challenge of pain identification, assessment, and management in people with dementia: a qualitative study. BJGP Open. 2020;4:bjgpopen20X101040

68. Liu JY. Exploring nursing assistants' roles in the process of pain management for cognitively impaired nursing home residents: a qualitative study. J Adv Nurs. 2014;70:1065-77.

69. Bullock L, Bedson J, Jordan JL, Bartlam B, Chew-Graham CA, Campbell P. Pain assessment and pain treatment for community-dwelling people with dementia: A systematic review and narrative synthesis. Int J Geriatr Psychiatry. 2019;34:807-21.

70. Gilmore-Bykovskyi AL, Bowers BJ. Understanding nurses' decisions to treat pain in nursing home residents with dementia. Res Gerontol Nurs. 2013;6:127-38.

71. Whybrow P, Moffatt S, Kay L, Thompson B, Aspray T, Duncan R. Assessing the need for arthritis training among paid carers in UK residential care homes: A focus group and interview study. Musculoskeletal Care. 2018;16:82-9.

72. Hunt LJ, Covinsky KE, Yaffe K, et al. Pain in community-dwelling older adults with dementia: results from the National Health and Aging Trends Study. J Am Geriatr Soc. 2015;63:1503-11.

73. Achterberg WP, Erdal A, Husebo BS, Kunz M, Lautenbacher S. Are chronic pain patients with dementia being undermedicated?. J Pain Res. 2021;14:431-9.

74. Roitto HM, Kautiainen H, Aalto UL, Öhman H, Laurila J, Pitkälä KH. Fourteen-year trends in the use of psychotropic medications, opioids, and other sedatives among institutionalized older people in Helsinki, Finland. J Am Med Dir Assoc. 2019;20:305–11.

FIGURE AND TABLE LEGENDS

Figure 1: PRISMA flow-chart reporting search results for Search 2

Table 1: Summary of trial registers which reported measuring pain in people with dementia

Table 2: Summary of included studies

Table 3: Best evidence synthesis of outcome measures used to assess pain in people with dementia against the COSMIN Risk of Bias checklist rating and level of evidence for the measurement property

Supplementary File 1: Search strategy (EMBASE example – optimised for other databases)

Supplementary File 2: The included pain assessment instruments examined with their original reference.

Supplementary File 3: Results of the COSMIN methodological quality assessment for each included study

Supplementary File 4: Psychometric results extracted by study

Figure 1: PRISMA flow-chart reporting search results for Search 2

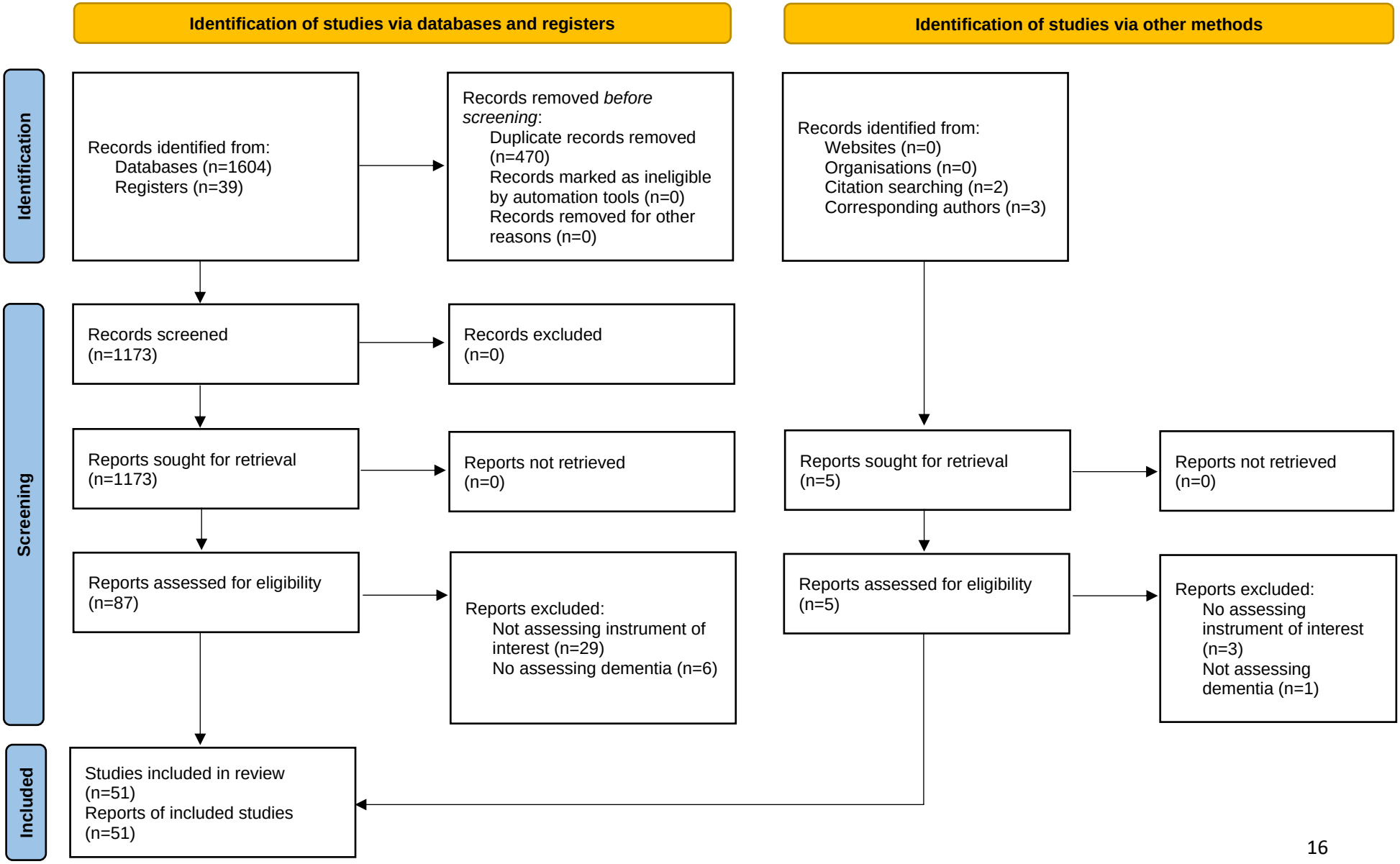

Table 1: Summary of trial registers which reported measuring pain in people with dementia

		Frequency	%
N		56	100
Date study commenced	2007-2011	2	3.6
	2012-2016	17	30.4
	2017-2021	37	66.0
Country of origin	Australia	1	1.8
	Belgium	1	1.8
	Canada	6	10.7
	China	1	1.8
	France	7	12.5
	Germany	1	1.8
	Italy	2	3.6
	Netherlands	2	3.6
	Norway	7	12.5
	Spain	2	3.6
	Switzerland	1	1.8
	Taiwan	3	5.4
	UK	3	5.4
	USA	19	33.9
Type of intention	Pharmacology agent	13	23.2
	Non-pharmacology intervention	43	76.8
Mean N (SD)		268.2 (576.1)	
Participant degree of cogitative impairment	Mild	11	19.6
	Mild-Moderate	10	17.9
	Mild-Severe	14	25.0
	Moderate-Severe	14	25.0
	Severe	7	12.5
Setting	Hospital	9	16.1
	Community-dwelling	21	37.5
	Care home	22	39.3
	Not stated	4	7.1
Mean follow-up period (SD)		26.2 (25.9)	
Pain Measure	Abbey Pain Scale	2	3.6
	ALGOPLUS	1	1.8
	Brief Pain Inventory	1	1.8
	Comfort Assessment in Dying with Dementia	1	1.8
	DOLOPLUS-2	1	1.8
	Edmonton Symptom Assessment Scale	2	3.6
	EQ-5D	5	8.9
	Facial Action Coding System	1	1.8
	GLOBAL PROMIS-10	1	1.8
	McGill Pain Map	1	1.8
	SF-36	3	5.4
	Medication use	2	3.6
	MOBID-2	9	16.1
	Self-reported (NRS/VAS Pain/Verbal Descriptor Scale/Thermometer)	9	16.1
	PACSLAC and PACSLAC-2	6	10.7
	Pain Assessment in Advanced Dementia (PAINAD)	9	16.1
	Philadelphia Geriatric Pain Intensity Scale Patient and Caregiver Responded	2	3.6

	Resident Assessment Index-Minimum Dataset	2	3.6
	Symptom Management - End of Life for Dementia	1	1.8
	WOMAC	1	1.8
	Checklist for non-verbal pain behavior	1	1.8

SD – standard deviation

Table 2: Summary of included studies

Study	Setting (hospital; home; care home)	Language	Severity of CI (mean)	Cohort Characteristics				Country of Origin	Study funding source	Measurement of pain
				N CI	Age (Mean)	Gender	Cognitive Diagnosis			
Abbey [35]	Care facility	English	Moderate to severe	61	Median : 83	40F/21M	Dementia	Australia	JH and JD Gunn Medical Research Foundation	Abbey Pain Scale
Akbarzadeh [50]	Care facility	Swedish	Not stated	48	>65	N/S	Dementia	Sweden	None declared	DOLOPLUS-2
Ando [25]	Hospital	Japanese	MMSE: 10	9	80	4F/5M	Dementia	Japan	None declared	DOLOPLUS-2
Ando [26]	Hospital	Japanese	MMSE: 10.9	19	84.5	15F/4M	Dementia	Japan	Okochi Fund at Yokufukai Geriatric Hospital	DOLOPLUS-2
Atee [39]	Care facility	English	PASCI score: 19.7	34	85.5	20F/14M	Dementia	Australia	Alzheimer's Australia Dementia Research Foundation	Abbey Pain Scale
Babicova [38]	Care facility	English	Moderate to severe Mean: 5.8	22	84.7	17F/5M	Dementia	UK	None declared	Abbey Pain Scale
Batalha [21]	Hospital	Portuguese	Not stated	99	82	68F/31M	Dementia	Portugal	None declared	PAINAD-P
Browne [30]	Care facility	English	CPS: 3.74	48	78.8	34F/19M	Dementia	Canada	AGE WELL Network of Centres of Excellence and the Canadian Institutes of Health Research	Facial Action Coding System PACSLAC-II
Büyükturan [16]	Hospital	Turkish	MMSE: 2.15	106	72.5	54F/52M	Dementia	Turkey	None declared	PAINAD-TR
Cantón-Habas [24]	Hospital	Spanish	GDS: 5-7	100	83.8	22F/78M	Dementia	Spain	Junta de Andalucía	PAINAD-S
Cantón-Habas [65]	Hospital and Care facility	Spanish	GDS: 5-7	75	84.4	59F/16M	Dementia	Spain	Junta de Andalucía	PAINAD-S
Chan [55]	Care facility	English	MMSE: 5.35	124	83.9	88F/36M	Dementia	Canada	Alzheimer Society of Canada; Saskatchewan Health Research Foundation; University of Regina	PACSLAC-II

Chen [49]	Care facility	Chinese	MMSE: 7.46	304	79.9	129F/175M	Dementia	Taiwan	National Science Council, Taiwan	DOLOPLUS-2
Chen [51]	Care facility	Chinese	MMSE: 5.26	241	79.2	118F/123M	Dementia	Taiwan	National Science Council, Taiwan	DOLOPLUS-2
Cheung [53]	Care facility	English	MMSE:7.5	50	82.9	36F/14M	Dementia	New Zealand	None declared	PACSLAC
Costardi [20]	Hospital	Italian	MMSE: 16.4	20	82	16F/4M	Dementia	Italy	None declared	PAINAD-Italian
Ersek [44]	Care facility	English	CPS: 3.9	60	89.0	53F/7M	Dementia	USA	National Institute of Nursing Research, USA	CNPI; PAINAD
Ersek [42]	Care facility	English	Severe	326	83.2	225F/101M	Dementia	USA	National Institute of Nursing Research, USA	Iowa Pain Thermometer; CNPI
Feldt [19]	Hospital	English	MMSE: CI 12.2 nCI: 27.2	88	83.2	76F/12M	Dementia	USA	University of Minnesota, USA	CNPI
Fuchs-Lacelle [52]	Care facility	English	PFQ: 44.6	40	83.2	29F/11M	Dementia	Canada	Saskatchewan Health Research Foundation; Canadian Institutes of Health Research Career Investigator Award	PACSLAC
Hadjistavropoulos [31]	Care facility	English	CPS: 3.74	48	82.5	69F/36M	Dementia	Canada	AGE WELL Network of Centres of Excellence and the Canadian Institutes of Health Research	Facial Action Coding System PACSLAC-II
Herr [60]	Care facility	English	Moderate-severe	138	84	63F/75M	Dementia	USA	Department of Veterans Affairs, USA	MOBID
Holen [32]	Care facility	Norwegian	MMSE Median: 9	59	Median: 82	47F/12M	Dementia	Norway	The Research Council of Norway	DOLOPLUS-2
Holen [63]	Care facility and hospital	Norwegian	Median MMSE: 10	73	84	54F/19M	Dementia	Norway	The Research Council of Norway	DOLOPLUS-2
Husebo [56]	Care facility	Norwegian	MMSE: 4.3	26	87.0	23F/3M	Dementia	Norway	The Research Council of Norway; Kavli's Research Center for Dementia.	MOBID
Husebo [58]	Care facility	Norwegian	MMSE: 4.3	26	87.0	23F/3M	Dementia	Norway	The Research Council of Norway; Kavli's Research Center for Dementia.	MOBID

Husebo [57]	Care facility	Norwegian	MMSE: 2.4	77	84.1	61F/16M	Dementia	Norway	The Research Council of Norway; Kavli's Research Center for Dementia.	MOBID-2
Husebo [59]	Care facility	Norwegian	MMSE: 8.1	203	85.4	149F/54 M	Dementia	Norway	The Research Council of Norway; Kavli's Research Center for Dementia.	MOBID-2
Kaasalainen [54]	Care facility	English	Not stated	338	82.8	216F/12 2M	Dementia	Canada	Canadian Institutes of Health Research	PACSLAC
Kunz [18]	Hospital	German	MMSE CI: 16.3 Healthy: 29.5	42	76.7	22F/20M	Dementia	Germany	Deutsche Forschungsgemeinschaft	Facial Action Coding System
Lautenbacher [64]	Hospital and Care facility	German	MMSE CI: 17.0 Healthy: 29.1	40	>65	N/S	Dementia	Germany	European Cooperation in the field of Scientific and Technical Research program; Oberfranken-Stiftung	Facial Action Coding System using the PAIC-FACE-SCALE
Leong [45]	Care facility	Chinese	CPS: 3.9	88	79.6	54F/34M	Dementia	Singapor e	Tan Tock Seng Hospital	Self-reported pain, PAINAD
Lin [47]	Care facility	Chinese	MMSE: 3.20	61	76.3	29F/32M	Dementia	China	National Science Council, Taiwan	PAINAD-C
Liu [40]	Care facility	Chinese	MMSE: CI: 9.97 nCI: 22.71	124	87.1	120F/4M	Dementia	Hong Kong	None declared	PAINAD, PACSLAC, Abbey Pain Scale
Lorenzet [66]	Not stated	Portuguese	Not stated	N/S	N/S	N/S	Not stated	Brazil	None declared	PACSLAC
Neville [43]	Care facility	English	Moderate to severe	126	85.2	104F/22 M	Dementia	Australia	University of Queensland	Abbey Pain Scale; DOLOPLUS 2; CNPI
Nygaard [41]	Care facility	Norwegian	SPMQ: 46 missing 2 answers	46	84.7	29F/17M	Dementia (89%)	Norway	Lions Foundation	CNPI
Parmelee [33]	Care facility	English	386 mild-severe CI	758	83.3	531F/22 7M	Dementia	USA	None declared	Self-Reported Pain and Pain Thermometer

Pateux [28]	Hospital	French	MMSE: 18.0	180	83.7	133F/47 M	Dementia	France	University Hospital of Geneva	Verbal rating scale; DOLOPLUS-2
Pautex [27]	Hospital	French	MMSE: 17.8	160	85.5	114F/46 M	Dementia	France	University Hospital of Geneva	Verbal rating scale; Faces Pain Scale
Pinto [23]	Hospital	Portuguese	N/S	66	Median : 87	44F/22M	Dementia	Brazil	None declared	PAINAD-Br
Rat [29]	Hospital	French	N/S	349	81.6	214F/13 5M	Dementia	France	CNP Foundation; Laboratoires Grünenthal France	Algoplus
Sefcik [61]	Care facility	English	N/S	197	84	95F/102 M	Dementia	USA	None declared	MOBID
Takai [36]	Care facility	Japanese	MMSE: 9.1	171	85.4	142F/29 M	Dementia	Japan	Kinuko Takasaki Gerontological Nursing Grant	Abbey Pain Scale-Japanese
Thé [17]	Care facility	Portuguese	N/S	50	87.8	39F/11M	Dementia	Brazil	None declared	PACSLAC
Torvik [48]	Care facility	Norwegian	MMSE: 0	77	86	58F/19M	Dementia	Norway	None declared	DOLOPLUS-2
Valera [22]	Hospital	Portuguese	N/S	27	81.8	19F/8M	Dementia	Brazil	São Paulo – FAPESP; Brazilian National Counsil of Scientific and Technological Development – CNPq	PAINAD-Br
Van Iersel [37]	Care facility	Dutch	N/S	157	85	122F/35 M	N/S	Belgium	None declared	Abbey Pain Scale-Dutch; PAINAD-Dutch
Weiner [34]	Care facility	English	N/S	115	Median : 81	51F/64M	Dementia	USA	National Institute of Health, USA; Arthritis Foundation, USA	Self-Reported Pain and Pain Thermometer
Zare [62]	Care facility and home	Persian	Mild-severe	100	87.3	71F/29M	Dementia	Iran	Kashan University of Medical Sciences	P-DOLOPLUS-2; PACSLAC-2-IR
Zwakhalen [46]	Care facility	Dutch	Mild-severe	128	82.4	100F/28 M	Dementia	Netherlands	The Netherlands Organization for Scientific Research	PAINAD, PACSLAC, DOLOPLUS-2

CI – Cognitively Impaired: CNPI - Checklist of Nonverbal Pain Indicators; CPS - Cognitive Performance Scale; F – Female; GDS - Global Deterioration Score; M – Male; MMSE: Mini-Mental State Examination; N/S – Not stated; nCI – not cognitively impaired; PASCI - Psychogeriatric Assessment Scale – Cognitive impairment; SPMQ - Short-Portable Mental Status Questionnaire

Table 3: Best evidence synthesis of outcome measures used to assess pain in people with dementia against the COSMIN Risk of Bias checklist rating and level of evidence for the measurement property

Measurement property	Frequency Assessed (N; Study)	COSMIN risk of bias checklist rating (N)				Level of evidence for measurement property	Overall rating
		Very good	Adequate	Doubtful	Inadequate		
Facial Action Coding System							
PROM Development	0 (0)						**MODERATE**
Construct validity	182 (3)		182				
Structural validity	40 (1)		40				
Internal consistency	0 (0)						
Cross-cultural validity	0 (0)						
Reliability	143 (1)		143				
Measurement error	0 (0)						
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	0 (0)						
PACSLAC-II							
PROM Development	0 (0)						**MODERATE**
Construct validity	224 (2)		224				
Structural validity	0 (0)						
Internal consistency	124 (1)	124					
Cross-cultural validity	0 (0)						
Reliability	267 (1)	267					
Measurement error	0 (0)						
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	124 (1)				124		
PACSLAC							
PROM Development	40 (1)	40					**MODERATE**
Construct validity	556 (4)	128	438				
Structural validity	124 (1)			124			
Internal consistency	342 (4)		342				
Cross-cultural validity	0 (1)			0			

Reliability	690 (5)	128	562				
Measurement error	0 (0)						
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	338 (1)		338				
Self-Reported Pain and Pain Thermometer							
PROM Development	88 (1)		88				**UNKNOWN**
Construct validity	882 (4)	702	180				
Structural validity	0 (0)						
Internal consistency	758 (1)				758		
Cross-cultural validity	0 (0)						
Reliability	1033 (3)	873	160				
Measurement error	0 (0)						
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	338 (1)				338		
Abbey Pain Scale							
PROM Development	61 (1)				61		**LIMITED**
Construct validity	571 (6)		517				
Structural validity	124 (1)			124			
Internal consistency	504 (5)	126			378		
Cross-cultural validity	335 (2)				335		
Reliability	313 (4)	313					
Measurement error	0 (0)						
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	56 (2)				56		
PAINAD							
PROM Development	0 (0)						**LIMITED**
Construct validity	858 (9)		858				
Structural validity	456 (5)	230	61	66	99		
Internal consistency	658 (8)	638					
Cross-cultural validity	430 (6)				430		
Reliability	764 (9)	764					
Measurement error	0 (0)						

Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	61 (1)				61		
Checklist of Nonverbal Pain Indicators							
PROM Development	0 (0)						**MODERATE**
Construct validity	757 (6)		757				
Structural validity	0 (0)						
Internal consistency	261 (3)	186			75		
Cross-cultural validity	0 (0)						
Reliability	232 (3)	320					
Measurement error	0 (0)						
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	0 (0)						
DOLOPLUS-2							
PROM Development	0 (0)						**MODERATE**
Construct validity	1036 (8)		1036				
Structural validity	752 (5)	341		411			
Internal consistency	672 (5)	274					
Cross-cultural validity	409 (4)	341			68		
Reliability	901 (7)	901					
Measurement error	0 (0)		100				
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	19 (1)				19		
Algoplus							
PROM Development	249 (1)	249					**STRONG**
Construct validity	249 (1)	249					
Structural validity	0 (0)						
Internal consistency	249 (1)	249					
Cross-cultural validity	0 (0)						
Reliability	249 (1)	249					
Measurement error	0 (0)						
Criterion validity	0 (0)						
Content validity	0 (0)						

Responsiveness	249 (1)	249					
MOBID							
PROM Development	26 (1)		26				**MODERATE**
Construct validity	335 (2)		335				
Structural validity	0 (0)						
Internal consistency	361 (3)		36		197		
Cross-cultural validity	0 (0)						
Reliability	52 (2)	52					
Measurement error	0 (0)						
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	0 (0)						
MOBID-2							
PROM Development	77 (1)		77				**MODERATE**
Construct validity	77 (1)		77				
Structural validity	0 (0)						
Internal consistency	77 (1)	77					
Cross-cultural validity	0 (0)						
Reliability	280 (2)	280					
Measurement error	203 (1)		203				
Criterion validity	0 (0)						
Content validity	0 (0)						
Responsiveness	203 (1)				203		

Strong; Moderate; Limited; Unknown

Supplementary File 1: Search strategy (EMBASE example – optimised for other databases)

1. exp Pain/
2. exp Pain Threshold/
3. exp Pain Perception/
4. exp Myalgia/
5. exp Neuralgia/
6. exp Acute Pain/
7. exp Chronic Pain/
8. exp Hyperalgesia/
9. exp Neuritis/
10. exp Paresthesia/
11. (pain or discomfort or allodynia, or neuritis or neuropathy or myalgia or neuralgia or hyperalgesia or paresthesia or soreness or ache* or dys?sthesia or Nocicepti*).ti,ab
12. OR/1-11
13. (aged or elder* or seniors or (old* adj2 (people or person* or patient* or men or women))).mp.
14. exp dementia/
15. exp Alzheimer Disease/
16. exp Cognition Disorders/
17. cognitive impairment.mp.
18. Cognitive function*.mp.
19. (alzheimer* or dement* or "Frontotemporal lobar degeneration" or "Frontotemporal dement*" or Huntington or "Lewy Body disease").tw.
20. OR/14-19
21. AND/13,20
22. exp Pain Measurement/
23. exp psychometrics/
24. exp Symptom Assessment/
25. exp Self Report/
26. (assessment or self report or identification or recognition or detection or evaluation or appraisal or rating).ti,ab
27. (tool* or test*).ti,ab
28. (instrumentation or "validation studies" or "comparative study" or psychometr*[tiab] or clinimetr* or clinometr* or outcome assessment (health care)).ti,ab
29. (“outcome assessment" or "outcome measure*" or "observer variation" or "observer variation" or "health status indicators" or "reproducibility of results" or reproducib* or "discriminant analysis").ti,ab
30. (reliab* or unreliab* or valid* or "coefficient of variation" or coefficient or "internal consistency" or (cronbach* AND (alpha or alphas)) or (item and (correlation* or selection* or reduction*).ti,ab
31. (agreement or precision or imprecision or "precise values").tw
32. (test-retest or (test and retest) or (reliab* and (test or retest)).ti,ab
33. (interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intra-observer or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant).ti,ab
34. (Kappa or kappa's or kappas).ti,ab
35. (repeatab* or ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti,ab
36. (generaliza* or generalisa* or concordance or (intraclass and correlation*) or discriminative or "known group" or "factor analysis" or "factor analyses" or "factor structure" or "factor structures" or dimension*).ti,ab
37. (subscale* or (multitrait and scaling and (analysis or analyses)) or "item discriminant" or "interscale correlation*" or error or errors or "individual variability" or "interval variability" or "rate variability" or (variability AND (analysis OR values)).ti,ab
38. (uncertainty and (measurement or measuring)) or "standard error of measurement" or sensitiv*).t,ab
39. (responsive* or (limit and detection) or "minimal detectable concentration" or interpretab* or ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)) or (small* and (real or detectable) and (change or difference)) or "meaningful change").ti,ab
40. (“ceiling effect" or "floor effect").ti,ab
41. ("Item response model" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank"or "cross-cultural equivalence").ti,ab
42. OR/18-37
43. AND/12,21,38

Supplementary File 2: The included pain assessment instruments examined with their original reference.

Pain Assessment Instrument	Original Citation
Abbey Pain Scale	Abbey J, Piller N, De Bellis A, Esterman A, Parker D, Giles L, Lowcay B. The Abbey pain scale: a 1-minute numerical indicator for people with end-stage dementia. Int J Palliat Nurs. 2004;10:6-13.
ALGOPLUS	Rat P, Jouve E, Pickering G, Donnarel L, Nguyen L, Michel M, Capriz-Ribière F, Lefebvre-Chapiro S, Gauquelin F, Bonin-Guillaume S. Validation of an acute pain-behavior scale for older persons with inability to communicate verbally: Algoplus. Eur J Pain. 2011;15:198.e1-198.e10
Checklist for non-verbal pain behavior (CNPI)	Feldt KS. The checklist of nonverbal pain indicators (CNPI). Pain Manag Nurs. 2000;1:13-21.
DOLOPLUS-2	Wary B, collectief Doloplus: Doloplus-2, une échelle pour évaluer la douleur. Soins Gérontologie. 1999;19:25-7. Lefebre-Chapiro L, Doloplus group: The Doloplus 2 scale-evaluating pain in the elderly. European Journal of Palliative Care. 2001;8:191-4.
Facial Action Coding System	Ekman P, Friesen W. Facial Action Coding System: a technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto; 1978.
MOBID-2	Husebo BS, Strand LI, Moe-Nilssen R, Husebo SB, Ljunggren AE. Pain in older persons with severe dementia. Psychometric properties of the Mobilization-Observation-Behaviour-Intensity-Dementia (MOBID-2) Pain Scale in a clinical setting. Scand J Caring Sci. 2010;24:380-91.
Pain Assessment in Advanced Dementia (PAINAD)	Warden V, Hurley AC, Volicer L. Development and psychometric evaluation of the Pain Assessment in Advanced Dementia (PAINAD) scale. J Am Med Dir Assoc. 2003;4:9-15.
PACSLAC/PACSLAC-2	Fuchs-Lacelle S, Hadjistavropoulos HD: Development and preliminary validation of the Pain Assessment Checklist for Seniors with Limited Ability to Communicate (PACSLAC). Pain Management Nursing. 2004;1:37-49. Chan S, Hadjistavropoulos T, Williams J, Lints-Martindale A. Evidence-based development and initial validation of the pain assessment checklist for seniors with limited ability to communicate-II (PACSLAC-II). Clin J Pain. 2014;30:816-24.
Self-reported pain through the NRS or VAS/thermometer or Philadelphia Geriatric Pain Intensity Scale	Parmelee PA, Smith B, Katz I. Pain complaints and cognitive status among elderly institution residents. J Am Geriatr Soc. 1993;41:517-22.

Supplementary File 3: Results of the COSMIN methodological quality assessment for each included study

Study	PROM Development	Internal consistency	Reliability	Measure error	Construct Validity	Structural validity	Hypothese testing	Cross-cultural validity/	Criterion validity	Responsiveness	Interpret.
Abbey [35]											
Akbarzadeh [50]											
Ando [25]											
Ando [26]											
Atee [39]											
Babicova [38]											
Batalha [21]											
Browne [30]											
Büyükturan [16]											
Cantón-Habas [24]											
Cantón-Habas [65]											
Chan [55]											
Chen [49]											
Chen [51]											
Cheung [53]											
Costardi [20]											
Ersek [44]											
Ersek [42]											
Feldt [19]											
Fuchs-Lacelle [52]											
Hadjistavropoulos [31]											
Herr [60]											
Holen [32]											
Holen [63]											
Husebo [56]											
Husebo [58]											
Husebo [57]											
Husebo [59]											
Kaasalainen [54]											

Kunz [18]											
Lautenbacher [64]											
Leong [45]											
Lin [47]											
Liu [40]											
Lorenzet [66]											
Neville [43]											
Nygaard [41]											
Parmelee [33]											
Pateux [28]											
Pautex [27]											
Pinto [23]											
Rat [29]											
Sefcik [61]											
Takai [36]											
Thé [17]											
Torvik [48]											
Valera [22]											
Van Iersel [37]											
Weiner [34]											
Zare [62]											
Zwakhalen [46]											

Rating: VG - Very good; adequate; inadequate; doubtful; not reported

Supplementary File 4: Psychometric results extracted by study

	Construct validity	Structural validity	Reliability (ICC)	Internal consistency (Cronbach)	Responsiveness (to treatment or rest)	Measurement error
Abbey [35]	R=0.586; P<0.001			0.74	P<0.001	
Akbarzadeh [50]	R=0.698; P=<0.001 to UAB	36.9% variance explained	Inter: 0.90			
Ando [25]			Inter: 0.90			
Ando [26]					P<0.001	
Atee [39]	R=0.91 with PainChek (p<0.001)		Inter: 0.86 Intra: 0.90	0.95	p<0.001	
Babicova [38]	R=0.82 with PainChek (p<0.001)		Inter: 0.72 intra: 0.68	0.81	p<0.001	
Batalha [21]		61.1% variance explain	Intra:0.89	0.84		
Browne [30]			Inter: 0.94 Inter: 0.86			
Büyükturan [16]	CVI: 0.84 P<0.001 to nurse VAS	68.9% variance explained	Intra: 0.81			
Cantón-Habas [24]	CVI: 0.875 P<0.02 to medication use	62.5% variance explained	Inter: 0.94 Intra: 0.80-0.83	0.76		
Cantón-Habas [65]	P<0.01 to sTNF-RII and sIgA pain biomarkers	46.5% variance explained				
Chan [55]	R=0.68; p<0.01 to PACSLAC-II; p<0.01 to CNPI; R=0.79; p<0.01 to PACSLAC-II		Inter: 0.75-0.97 intra: 0.88-0.90 Inter: 0.63	0.69-0.80 0.74-0.77	p<0.01	
Chen [49]		70.4% variance explained	Inter: 0.35			
Chen [51]		65% variance explained	Inter: 0.81	0.74		
Cheung [53]			Inter: R=0.83			
Costardi [20]	R=0.65; p=0.008		Inter: R=0.87 Intra: R=0.88	0.74		
Ersek [44]	R=0.48; p<0.001; R=0.41; P<0.05		Intra:0.80 Inter:0.04 Intra:0.65	0.72; 0.90		

			Inter:0.25			
Ersek [42]	P<0.001					
Feldt [19]	R=0.46 (p<0.01)		Intra: 0.60 Inter: 0.86-0.90	0.64		
Fuchs-Lacelle [52]	R=0.54; p<0.001 to global intensity rating			0.82-0.87		
Hadjistavropoulos [31]	R=0.542; p<0.01; R=0.463; p<0.01		Inter: R=0.76; Inter: R=0.89			
Herr [60]	R=0.54 P<0.001 to caregiver NRS			0.83		
Holen [32]		68% variance explained				
Holen [63]	R2=0.023		Intra: 0.74 Inter: 0.77			
Husebo [56]			Inter: 0.86	0.86-0.90		
Husebo [58]			Intra: 0.79-0.92 Inter: 0.86-0.97			
Husebo [57]	R2=0.61; P<0.01 to caregiver NRS		Inter: 0.94 Intra: 0.92	0.82-0.84		
Husebo [59]			Intra: 0.852		p<0.001 MCID: 3 points	SEM: 1.4
Kaasalainen [54]	p<0.01 with NRS; p<0.01 with PACSLAC		Inter: 0.87		p<0.001; p=0.03	
Kunz [18]	P<0.001					
Lautenbacher [64]	R^2=0.116 P<0.001	P=0.006 to p<0.001				
Leong [45]	R=0.304; p<0.01; R=0.842; p<0.001					
Lin [47]		62.5 variance explained	Intra: 0.71 Inter: 0.84	0.55	P<0.001	
Liu [40]	P<0.01	Cronbach: 0.75 Cronbach: 0.76	Inter: 0.90 Inter: 0.88 Inter: 0.82	0.73 0.72 0.77		
Neville [43]	P<0.01		Intra:0.68 Inter:0.75	0.74; 0.76; 0.86		

			Intra:0.71 Inter:0.73 Intra:0.56 Inter:0.59			
Nygaard [41]	R=0.88 (p<0.001)					
Parmelee [33]			Intra: 0.84	0.84		
Pateux [27]	R2=0.46 P<0.01 to DOLOPLUS; R2=0.46 P<0.01 to VAS self-assess					
Pateux [28]			Inter: 0.97 Intra: 0.71-0.80			
Pinto [23]			Inter: 0.79	0.65		
Rat [29]	R2=0.81 (p<0.001) to VAS Pain; R2=0.81 (p<0.001) to Algoplus		Inter: 0.812; Inter: 0.43-0.80	KR-20: 0.712	P<0.001	
Sefcik [61]	R2=0.51 P<0.001 to caregiver NRS			0.83		
Thé [17]	R=0.64; P<0.001 to caregiver VAS		Inter: 0.85 Inter: 0.64	0.827		
Takai [36]	R=0.49 P<0.01		Inter: 0.82 Intra: 0.66	0.65		
Torvik [48]	P=0.01 agree to nursing VAS			0.71		
Weiner [34]	R=0.95		Intra: 0.71-0.85			
Zare [62]		76.14% variance explained	Inter: 0.86	0.950		SEM: ± 1.759
Zwakhalen [46]	R=0.81 to VAS nurse (p<0.01)		Inter: 0.81 Intra: 0.89	0.72		
Zwakhalen [46]	R=0.72 to VAS nurse (p<0.01)		Inter: 0.96 Intra: 0.86	0.84		
Zwakhalen [46]	R=0.33 to VAS nurse (p<0.01)		Inter: 0.78 intra:0.85	0.74		

CVI: Content Validity Index; ICC – Intra-class correlation coefficient; MCID – minimally clinical important difference; NRS – numerical rating scale; SEM – standard error of mean; VAS – visual analogue scale