JCN Journal of Clinical Nursing

Journal of Clinical Nursing

REVIEW

A systematic review of observational pain assessment instruments for use with nonverbal intubated critically ill adult patients in the emergency department: an assessment of their suitability and psychometric properties

Wayne Varndell, Margaret Fry and Doug Elliott

Aim and objective. To examine the psychometric properties and suitability of the available observational pain instruments for potential use with nonverbal critically ill adult patients in the emergency department.

Background. In the emergency department, assessing pain in critically ill patients is challenging, especially those unable to communicate the presence of pain. Critically ill patients are commonly unable to verbally communicate pain due to altered oral communication (e.g. endotracheal intubation) and/or diminished consciousness (e.g. sedation, delirium), placing them at great risk of inadequate pain management. Over half of intensive care critically ill intubated patients experience moderate-to-severe pain whilst intubated and mechanically ventilated.

Design. Systematic review.

Data sources. The CINAHL, EMBASE, MEDLINE, ProQuest databases, and the Cochrane Library and the National Institute of Clinical Excellence were also searched from their date of inception to April 2016, with no language restrictions applied.

Review method. Studies were identified using predetermined inclusion criteria. Data were extracted and summarised and underwent evaluation using published classification of psychometric tests for consistency of interpretation.

Results. Twenty-six studies evaluating five observational pain assessment instruments that had been used with critically ill intubated patients were identified. All five instruments included behavioural indicators, with two including physiologic indicators. All five instruments have undergone validity and reliability testing involving nonverbal critically ill intubated patients, three were examined for feasibility, and one instrument underwent sensitivity and specificity testing. None have been tested within the emergency department with nonverbal critically ill intubated adult patients.

Conclusion. The use of an appropriate and valid observational pain assessment instrument is fundamental to detecting and optimising pain management in nonverbal critically ill intubated patients in the emergency department. Of the observational

What does this paper contribute to the wider global clinical community?

- Care of the critically ill intubated patients is traditionally associated with intensive care, yet it increasingly happens in the emergency department (ED) for prolonged periods of time and strongly relies upon the knowledge, skills and expertise of emergency nurses, including the detection and management of acute pain.
- Despite the frequency of patients presenting to the ED complaining of pain, pain is often poorly detected and managed within the emergency department, a situation potential worse for nonverbal critically ill intubated patients.
- Critically ill intubated patients frequently experience moderate-to-severe acute pain; the use of an appropriate observational pain assessment instrument is fundamental to improving pain assessment and management for critically ill intubated adult patients in the ED.

Authors: *Wayne Varndell*, MN, BSc, GradDip, PGCert, PhD Candidate, Prince of Wales Hospital Emergency Department, Randwick, NSW and Faculty of Health, University of Technology Sydney, Broadway, NSW; *Margaret Fry*, PhD, NP, Professor, Faculty of Health, University of Technology Sydney, Broadway, NSW; *Doug Elliott*, PhD, BAppSc, MAppSci, Professor, Faculty of Health, University of Technology Sydney, Broadway, NSW, Australia

Correspondence: Wayne Varndell, PhD Candidate, Prince of Wales Hospital Emergency Department, Barker Street, Randwick, NSW 2031, Australia. Telephone: +02 9382 3910.

E-mail: waynevarndell@mac.com

© 2016 John Wiley & Sons Ltd

pain assessment instruments reviewed, the Critical-Care Pain Observation Tool was identified as most appropriate for testing in a prospective trial in an emergency department setting.

Key words: critical care, emergency department, pain, pain assessment, pain rating scales

Accepted for publication: 22 September 2016

Introduction

Nonverbal critically ill intubated adult patients unable to communicate in the emergency department have complex needs. While adequate pain relief in the emergency department (ED) is paramount in optimising comfort, it has been widely reported that over half of this patient cohort experience moderate-to-severe pain whilst intubated and mechanically ventilated (Desbiens *et al.* 1996, Nelson *et al.* 2001, 2004, Puntillo *et al.* 2001, Li & Puntillo 2006, Ma *et al.* 2010). Pain may be multifactorial in terms of presentation (e.g. trauma), but also iatrogenic. Critically ill patients are exposed to multiple sources of pain such as repositioning, physical examination (e.g. palpation, assessing limb movement), endotracheal intubation, mechanical ventilation, insertion of central venous catheters and chest drains, which commonly occur during resuscitation and stabilisation in the ED (Weir & O'Neill 2008, Varndell *et al.* 2015b). However, pain remains undetected, underestimated and poorly managed, particularly in critically ill intubated patients (Thomas 2013, Batiha 2014, Clukey *et al.* 2014).

Pain is a subjective, complex and multidimensional concept. It is broadly defined as an unpleasant sensory experience associated with actual or potential tissue damage (International Association for the Study of Pain, 2006). In the ED, critically ill intubated patients commonly require intravenous analgesia, often in combination with sedation to prevent suffering, adverse physiologic and psychological effects (Schug *et al.* 2015). Quality pain management begins with effective assessment of pain. International guidelines recommend pain must be adequately assessed in all patients, with appropriate and timely analgesia provided at the point of assessment (Curtis & Morrell 2006, Erstad *et al.* 2009, Schug *et al.* 2015).

While care of the critically ill patients traditionally occurs in intensive care units, it increasingly happens in ED for extended time periods (Rose *et al.* 2012), placing substantial demand on departmental resources (O'Connor *et al.* 2009). Assessing and monitoring pain in intubated critically ill patients is challenging. Critically ill intubated patients are commonly unable to verbally report pain due to altered oral communication (e.g. endotracheal intubation) and/or altered levels of consciousness (e.g. sedation, delirium), placing them at great risk of inadequate pain management (Jacobi *et al.* 2002, Puntillo *et al.* 2004, 2009, Varndell *et al.* 2015a). In the absence of self-reporting, pain behaviours such as facial grimacing, crying and ventilator compliance form the basis for identifying and evaluating the patient's degree of pain (Herr *et al.* 2006). Pain behaviours may only be fleetingly present, requiring emergency nurses to remain vigilant for physical changes, while assessing and responding to haemodynamic changes in the patient's condition.

Nonverbal critically ill intubated patients require close observation and monitoring during treatment of their underlying condition, with titration of analgesia, sedation and ventilator settings (Aitken *et al.* 2009), now a role increasingly undertaken by emergency nurses (Varndell *et al.* 2015c). In the absence of communicating pain levels, observational assessment instruments with unidimensional measures with multiple domains or multidimensional measures are preferred to evaluate acute pain in nonverbal critically ill patients (Herr *et al.* 2011). A unidimensional objective measure such as a behavioural scale may use a single domain (e.g. facial expression) or several domains (e.g. facial expression, body movements) to evaluate a person's responses to pain. A multidimensional objective measure evaluates two or more pain dimensions (e.g. behaviours, physiologic responses) and has several domains within each dimension. While a range of nonverbal pain assessment instruments are available to inform and guide nurse decision-making, there is no evidence of their use or evaluation in the ED setting.

The review

Aim

This systematic review summarises the psychometric properties of instruments assessing pain in nonverbal critically

© 2016 John Wiley & Sons Ltd

ill intubated adult patients. It addresses the following research questions:

1 What observational pain assessment instruments are available for use by emergency nurses in assessing pain in nonverbal critically ill intubated adult patients?
2 To what extent have the instruments been tested for validity, reliability, responsiveness and feasibility?
3 What observational pain assessment instruments would be suitable to be evaluated prospectively in assessing pain in nonverbal critically ill intubated patients in an ED setting?

Design

This was a systematic review of published research on the suitability and measurement properties of pain assessment instruments for nonverbal critically ill adult patients and was guided by The Cochrane Collaboration systematic review method (Higgins & Green 2011). The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) reporting guideline was used to present the review process (Fig. 1) and guide the reporting of study methods and results (Moher *et al.* 2009). Synthesis of the results focussed on evaluating the psychometric properties of instruments using published classification of psychometric tests for consistency of interpretation and appraising the characteristics and methodological quality of studies (Streiner *et al.* 2015).

Search methods

A systematic literature search was conducted using the following databases: CINAHL, EMBASE, MEDLINE and ProQuest. The Cochrane Library and the National Institute of

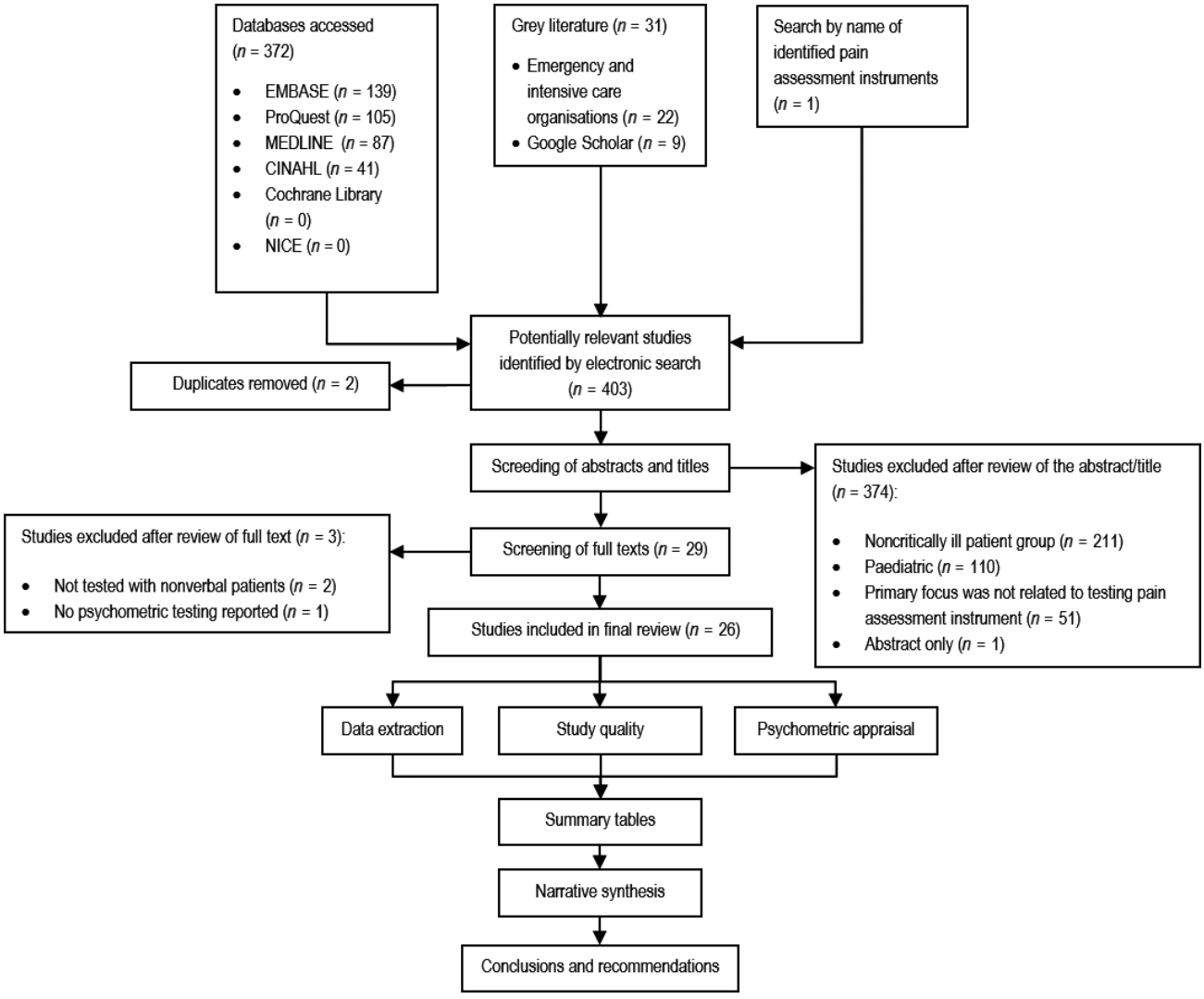

Figure 1 PRISMA flowchart of the search and systematic review process.

© 2016 John Wiley & Sons Ltd

Clinical Excellence were also searched. The review was supplemented with a manual search of the relevant grey literature for the purposes of identifying additional scholarly research. The database search was limited to scholarly articles or peer-reviewed journals. No language restrictions were applied. Search strategies were customised for each database due to the different range of search interfaces. Several search terms were used to identify potential studies concerning the assessment of pain in critically ill nonverbal adult (≥16 years) patients: 'pain assessment', 'pain scale AND non-verbal OR nonverbal', 'pain AND management', 'pain assessment AND intubated', 'pain assessment AND emergency OR critical care', 'pain assessment AND emergency nurse OR critical care nurse' and 'measuring pain AND critical care'. The reference lists of selected studies were hand-searched. Names of instruments found were added as keywords in an iterative process to further identify additional studies.

Exclusion criteria

Studies meeting the following criteria were excluded from the review:

1 did not mention the background of participants,
2 abstracts, duplicate or incomplete reports,
3 did not report results for each study variable.

Eligibility criteria

Studies meeting all the following criteria were included in the review:

1 assessed pain in critically ill adults,
2 described a unidimensional measure with multiple domains or multidimensional pain measure,
3 the instrument was used with nonverbal patients,
4 the instrument had undergone psychometric evaluation.

Search outcome

The initial search yielded 403 references after duplicates were removed. After a preliminary review of the titles and abstracts, 375 (93%) studies were excluded because the eligibility criteria were not fully met. A total of 29 studies were retained for full-text review, subsequent to which a further three studies were excluded as they did not provide sufficient evidence of meeting the third (Mateo & Krenzischek 1992, Klien *et al.* 2010) or fourth inclusion criteria (Puntillo *et al.* 2004). Therefore, 26 studies representing five observation pain assessment instruments were included in the systematic review and underwent quality appraisal and examination of their psychometric properties: the Behavioural Pain Scale (BPS); Critical-Care Pain Observation Tool (CPOT); Faces, Legs, Activity, Cry and Consolability Scale (FLACC); Pain Assessment in Advanced Dementia (PAINAD); and the Nonverbal Pain Scale (NVPS).

Quality appraisal, data extraction and synthesis

The primary author comprehensively examined the selected studies using the Critical Appraisal Skills Program (CASP) checklist for cohort studies for quality and risk of bias (Critical Appraisal Skills Program UK, 2011). The checklist consists of 12 criteria divided across three sections, answers to which can range from 'Yes', 'No' and 'Can't tell'. For the purposes of this review, 'Yes' was given a value of 1, with 'No' and 'Can't tell' a value of 0 (Tripathi & Dusing 2015) (Table 1).

The critical evaluation and deposition of articles was discussed with the other two authors (MF and DE), with consensus determining final inclusion. Data extraction was conducted by the primary author and included: (1) name of the instrument(s) tested, country of origin and the authors;

Table 1 Critical Appraisal Skills Program checklist for cohort studies

Criteria, example
Clear aim
The study addressed a clearly focused issue
Acceptable sampling
Sample recruited in an acceptable manner, representative of a defined population
Exposure accurately measured
Procedure clearly articulated
Outcome accurately measured
Independent assessors, valid subjective and objective measures employed, patient's self-report
Confounding factor(s) identified
Confounding factors identified and accounted for in the study design/data analysis (e.g. level of sedation and delirium)
Follow-up sufficiently complete
Adequate exposure to and assessment under all planned interventions (e.g. nonpainful, painful stimuli and rest)
Results clearly described
Strength of exposure and outcomes, results are complete and supported by appropriate statistical testing
Reporting probability values
Confidence intervals, probability testing undertaken
Do you believe the results?
Appropriate study design and methods employed, risk of bias
Results applicable to local population
Subjects examined in the study are similar to the local clinical population
Results fit with other evidence
Results similar with other studies
Implications for practice
Results are robust and consistent with other studies

© 2016 John Wiley & Sons Ltd

(2) instrument domains, items and scoring system; (3) study design, setting and characteristics of/and sample size; (4) method of evaluation including statistical analyses employed, number of raters; (5) key findings, limitations and strengths. The extracted data were summarised and synthesised in narrative and descriptive numerical form. The remaining authors verified the accuracy of the extraction and synthesis by way of random audit.

Several appraisal methods have been developed for evaluating the validity of measurement tools, but none have been rigorously evaluated (Greenhalgh *et al.* 1998, Jerosch-Herold 2005, Terwee *et al.* 2007, 2012, Leung *et al.* 2012). Terminology and application of statistical tests used within the identified studies differed depending on the research paradigm used and the purpose of testing. As such, we focused on performing a comprehensive assessment of the psychometric properties of the scores obtained using the research use measures reported in each study and applied published classification of psychometric tests for consistency of interpretation (Streiner *et al.* 2015). Statistical meta-analysis was not possible due to the heterogeneity of study samples, methods and reporting of results.

Results

Study characteristics

Included studies were published between 2001–2016, from a wide range of countries: Canada (n = 8); USA (n = 8); the Netherlands (n = 3); Australia (n = 1); France (n = 1); North Africa (n = 1); South America (n = 1); Spain (n = 1); Sweden (n = 1); and Taiwan (n = 1). All studies reported using convenience sampling and mostly consisted of small numbers (mode 30, median 64, range 25–239), and largely conducted in a single intensive care setting (Payen *et al.* 2001, Odhner *et al.* 2003, Aïssaoui *et al.* 2005, Gélinas *et al.* 2006, 2009, Young *et al.* 2006, Gélinas & Johnston 2007, Ahlers *et al.* 2008, 2010, Kabes *et al.* 2009, Gélinas 2010, Marmo & Fowler 2010, Topolovec-Vranic *et al.* 2010, 2013, Voepel-Lewis *et al.* 2010, Chen *et al.* 2011, Damström *et al.* 2011, Gélinas *et al.* 2011, Vazquez *et al.* 2011, Linde *et al.* 2013, Chanques *et al.* 2014, Morete *et al.* 2014). Fourteen studies reported raters by professional group of which the majority were intensive care nurses (n = 383), compared to intensive care physicians (n = 12), observers/researchers (n = 6) or allied health (n = 1) (Payen *et al.* 2001, Aïssaoui *et al.* 2005, Young *et al.* 2006, Gélinas & Johnston 2007, Kabes *et al.* 2009, Ahlers *et al.* 2010, Marmo & Fowler 2010, Topolovec-Vranic *et al.* 2010, 2013, Chen *et al.* 2011, Damström *et al.* 2011, Paulson-Conger *et al.* 2011, Vazquez *et al.* 2011, Chanques *et al.* 2014, Morete *et al.* 2014).

Instruments identified

The review covered 26 studies that reported on five nonverbal pain assessment instruments: BPS, CPOT, FLACC, PAINAD and the NVPS. Of these, four (BPS, CPOT, PAINAD and NVPS) were unidimensional objective pain instruments, and one was a multidimensional instrument (FLACC). Domains commonly (n = 4; 80%) consisted of three items weighted 0 to 2 (FLACC, CPOT, PAINAD and NVPS) or four items weighted 1 to 4 (BPS).

All instruments assessed facial expression, with end-anchor points varying from relaxed (BPS and CPOT) or smiling (FLACC, PAINAD and NVPS) to grimacing (CPOT). Four (BPS, FLACC, CPOT and NVPS) of the five instruments assessed body movement and posturing domains, and three (FLACC, CPOT and NVPS) instruments contained both (Odhner *et al.* 2003, Gélinas *et al.* 2006, Voepel-Lewis *et al.* 2010). Patient verbal response was rated in three instruments (CPOT, FLACC and PAINAD). End-points consisted of no verbal response (PAINAD), talking in normal tone (CPOT) or not crying (FLACC) to crying (CPOT, FLACC and PAINAD). Two instruments assessed body movement by focusing on the arms (BPS) or legs (NVPS). Two of the five instruments measured patient compliance with ventilation with end-points of tolerating movements (BPS and CPOT) to fighting ventilator (CPOT) or unable to control ventilation (BPS). Four of the five instruments assessed body posture with end-points of relaxed (CPOT and PAINAD) or normal position (FLACC and NVPS) through to being rigid (CPOT and NVPS), hitting out (PAINAD) or restless (FLACC) (Table 2).

All instruments were tested for validity and reliability; however, there was limited feasibility (Payen *et al.* 2001, Odhner *et al.* 2003, Topolovec-Vranic *et al.* 2010, Gélinas *et al.* 2011) and responsiveness testing (Aïssaoui *et al.* 2005, Ahlers *et al.* 2010). Four studies examined convergent validity, comparing the test instrument to BPS (Juarez *et al.* 2010, Chanques *et al.* 2014), NVPS (Odhner *et al.* 2003) or CPOT (Paulson-Conger *et al.* 2011), and one study compared change in pain score following administration of analgesia (Voepel-Lewis *et al.* 2010). Seven studies examined concurrent validity by comparing scores with patient self-reports of pain (Gélinas *et al.* 2006, 2009, Gélinas & Johnston 2007, Topolovec-Vranic *et al.* 2010, 2013, Chen *et al.* 2011, Damström *et al.* 2011). One instrument (CPOT) underwent sensitivity and specificity testing (Gélinas & Johnston 2007, Gélinas *et al.* 2009) (Table 3).

© 2016 John Wiley & Sons Ltd

Table 2 Dimensions, domains and item weighting of selected objective pain assessment instruments

Instrument	Scoring method and total score	Behavioural dimension, domains						Physiologic dimension
		Facial expression	Body movement	Body posture	Verbal response	Ventilator compliance	Other	
Behavioural Pain Scale (BPS)	No. items: 12 Item weighting: 1–4 Total score: 3 (no pain present) to 12 (most pain present)	Relaxed (1) Partially tightened (2) Fully tightened (3) Grimacing (4)	*Upper limb* No movement (1) Partially bent (2) Fully bent with finger flexion (3) Permanently retracted (4)			Tolerating movement (1) Coughing but tolerating ventilation most of the time (2) Fighting ventilator (3) Unable to control ventilation (4)		
Critical-Care Pain Observation Tool (CPOT)	No. items: 15 Item weighting: 0–2 Total score: 0 (no pain present) to 8 (most pain present).	Relaxed, neutral (0) Tense (1) Grimacing (2)	Absence of movements (0) Protection (1) Restlessness (2)	Relaxed (0) Tense, rigid (1) Very tense or rigid (2)	Talking in normal tone or no sound (0) Sighing, moaning (1) Crying out, sobbing (2)	Tolerating ventilator or movement (0) Coughing but tolerating (1) Fighting ventilator (2)		
Faces, Legs, Activity, Cry and Consolability Scale (FLACC)	No. items: 15 Item weighting: 0–2 Total score: 0 (no pain present) to 10 (most pain present).	No particular expression or smile (0) Occasional grimace or frown, withdrawn, disinterested (1) Frequent to constant quivering chin, clenched jaw (2)	*Legs* Normal position or relaxed (0) Uneasy, restless, tense (1) Kicking, or legs drawn up (2)	Lying quietly, normal position, moves easily (0) Squirming, shifting back and forth, tense (1) Arched, rigid or jerking (2)	*Cry* No cry (awake or asleep)(0) Moans or whimpers, occasional complaint (1) Crying steadily, screams or sobs, frequent complaints (2)		*Consolability* Content, relaxed (0) Reassured by occasional touching, hugging or being talked to, distractible (1) Difficult to console or comfort (2)	
Pain Assessment in Advanced Dementia (PAINAD)	No. items: 12 Item weighting: 0–2 Total score: 0 (no pain present) to 10 (most pain present).	Smiling, or inexpressive (0) Sad, frightened, frown (1) Facial grimacing (2)		Relaxed (0) Tense, distressed, pacing, fidgeting (1) Rigid, fists clenched, knees	None (0) Occasional moan or groan, low-level speech with a negative or disapproving			*Breathing* Normal (0) Occasional laboured breathing, short period of hyperventilation

© 2016 John Wiley & Sons Ltd

Table 2 (continued)

Instrument	Scoring method and total score	Behavioural dimension, domains						Physiologic dimension
		Facial expression	Body movement	Body posture	Verbal response	Ventilator compliance	Other	
				pulled up, pulling or pushing away, striking out (2)	quality (1) Repeated troubled, calling out, loud moaning or groaning, crying (2)			(1) Noisy laboured breathing, long period of hyperventilation, Cheyne–stokes respiration (2)
Nonverbal Pain Scale (NVPS)	No. items: 15 Item weighting: 0–2 Total score: 0 (no pain present) to 10 (most pain present).	No particular expression or smile (0) Occasional grimace, tearing, frowning, wrinkled forehead (1) Frequent grimacing, tearing, frowning, wrinkled forehead (2)	Lying quietly, normal position (0) Seeking attention through movement or slow, cautious movement (1) Restless, excessive activity and/or withdrawal reflexes (2)	Lying quietly, no posturing or hands over areas of body (0) Splinting areas of the body, tense (1) Rigid, stiff (2)				*Physiologic I* Stable vital signs (no change in past 4 hours) (0) Change over past 4 hours in any of the following: SBP >20 mmHg, HR >20/minutes, RR > 10/minutes (1) Change over past 4 hours in any of the following: SBP > 30/mmHg, HR >25/minutes, RR > 20/minutes (2) *Physiologic II* Warm, dry skin (0) Dilated pupils, perspiring, flushing (1) Diaphoretic, pallor (2)

© 2016 John Wiley & Sons Ltd

Table 3 Methodological evaluation of nonverbal observational pain assessment instruments

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
Behavioural Pain Scale (BPS)	Payen *et al.* (2001), France	Descriptive design, convenience sample (n = 30 ICU patients), 269 assessments by 46 trained nurse raters	Construct validity: change in total score from baseline (prestimuli 3·0–3·3) to stimulation (non-nociception, 3·5 vs. nociception 4·9; p = <0·01) Discriminant validity: comparison between painful procedure scores (4·6–5·2) with nonpainful procedure scores (3·3–3·7), Factor analysis: first factor accounted for 55% of variance in pain expression with correlations between BPS domains observed: facial expression (ICC = 0·789), upper limb movement (ICC = 0·794) and compliance with ventilator (ICC = 0·632)	Internal consistency: Cronbach's α = 0·94 Inter-rater agreement: at rest, r^2 = 0·71 (p = <0·01), during procedure (painful stimuli), r^2 = 0·50 (p = <0·01)	Satisfaction survey: 28 of 34 questionnaires retuned by raters; 85·7% (24/28) were satisfied or very with ease of use; 25% (7/28) concerned about relative complexity of the BPS; 89·3% considered BPS effective in assessing pain	Not assessed
	Aïssaoui *et al.* (2005), Morocco	Descriptive design, convenience sample (n = 30 ICU patients), 360 assessments by 28 trained raters (n = 12 physicians, n = 16 nurses)	Construct validity: change in score (rest, 3·9 ± 1·1 vs. procedures, 6·8 ± 1·9; p = <0·001). Factor analysis: first factor accounted for 65% of variance in pain expression with correlations between BPS domains observed: facial expression (ICC = 0·90), upper limb movement (ICC = 0·85) and compliance with ventilator (ICC = 0·64)	Internal consistency: Cronbach's α = 0·72 Inter-rater agreement: facial expression, ICC = 0·91; upper limb movement, ICC = 0·90; compliance with ventilator, ICC = 0·89; and total score, ICC = 0·95; BPS score vs. HR r = −0·16 (p = 0·13), BPS vs. arterial BP r = −0·02 (p = 0·84)	Not assessed	Effect size: total score, 2·2–3·4.
	Young *et al.* (2006), Australia	Prospective, descriptive, repeated-measures design, convenience sample (n = 80 ICU patients), 175 assessments by 72 trained nurse raters (n = 126 deeply sedated, n = 46 lightly sedated)	Construct validity: change in mean score (rest, 3·36 vs. painful stimuli, 5·02; p < 0·001). Discriminant validity: change in mean score from rest (3·23) compared to after application of nonpainful procedure (3·23). The odds of an increase in BPS between pre- and postprocedure assessments was more than 25 times higher for repositioning compared with eye care (p < 0·001), after controlling for analgesics and sedatives	Inter-rater reliability: K = 0·80–83	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
	Ahlers *et al.* (2008), the Netherlands	Descriptive design, convenience sample (n = 113 ICU patients), 371 assessments	Moderate correlation (r = 55; p < 0·001) between NRS and BPS score	Inter-rater reliability: K = 0·67	Not assessed	Not assessed
	Ahlers *et al.* (2010), the Netherlands	Descriptive design, convenience sample (n = 80 ICU patients), 175 assessments (126 deep sedation; 49 conscious sedation) by 72 trained nurse raters	Construct validity: moderate (r = 0·67; p < 0·001) correlation between VRS and BPS scores. Discriminant validity: change in total score from rest to painful procedure for both sedated (3·4 vs. 5·1) and conscious sedated patients (3·8 vs. 5·4); no difference between rest and nonpainful procedures for both sedated (3·3 vs. 3·4) or conscious sedated patients (3·6 vs. 3·7)	Internal consistency: sedated and conscious sedated (Cronbach's α = 0·63 and 0·66, respectively). Inter-rater reliability: deeply sedated, total score K = 0·83, facial expression K = 0·80, upper limb movement K = 0·72 and ventilator compliance K = 0·62; conscious sedated, total score K = 0·80, facial expression K = 0·78, upper limb movement K = 0·67, ventilator compliance K = 0·61	Not assessed	Effect size: sedated vs. conscious sedated, total score (2·5–1·8), facial expression (3·6–2·4), ventilator compliance (1·4–0·9) and upper limb movement (0·7–0·5)
	Juarez *et al.* (2010), USA	Descriptive study, convenience sample (n = 124 surgical and n = 115 medical ICU patients), 968 assessments	Construct validity: change in score (rest, 3·73 vs. procedures, 5·41; p = <0·001). Convergent validity: comparison between BPS and NVPS total score (rest ρ = 0·69, procedure, ρ = 0·77; p < 0·001)	Internal consistency: overall, Cronbach's α = 0·70; surgical patients, Cronbach's α = 0·77, medical patients, Cronbach's α = 0·60. Inter-rater consistency: overall BPS score (rest 0·58, procedure, 0·68; p < 0·001); surgical subgroup (rest 0·68; procedure 0·73) and medical subgroup (rest 0·46; procedure 0·58)	Not assessed	Not assessed
	Chen *et al.* (2011), Taiwan	Descriptive design, convenience sample (n = 70 ICU patients), 700 assessments by two trained nurse raters	Construct validity: Change in BPS score from rest and during procedure (t = 2·28; p = 0·03)	Not assessed	Not assessed	Not assessed
	Morete *et al.* (2014), Brazil	Descriptive design, convenience sample (n = 100 ICU patients), seven raters (nurses n = 6, physiotherapist n = 1)		Internal consistency: overall, Cronbach's α = 0·501. Inter-rater agreement: 98·0% ventilator compliance; 85·1% upper limb movement; and 90·1% facial expression. Overall, ICC = 0·807	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
	Chanques *et al.* (2014), USA	Descriptive design, convenience sample (n = 30 ICU patients), 258 paired assessments, by 24 raters (n = 20 nurses, n = 4 observers)	Convergent validity: BPS score increases from baseline to procedure in line with CPOT and NVPS scores and 10 min^{-1} after procedure ($p < 0{\cdot}001$), nil (median) score differences noted but not significant ($p > 0{\cdot}05$)	Internal consistency: Overall, Cronbach's $\alpha = 0{\cdot}80$, intubated patients Cronbach's $\alpha = 0{\cdot}81$, nonintubated Cronbach's $\alpha = 0{\cdot}83$. Inter-rater reliability: Total score $K_w = 0{\cdot}81$; BPS domains, face $K_w = 0{\cdot}75$, ventilation/vocalisation $K_w = 0{\cdot}78$, and upper limbs $K_w = 61$	Not assessed	Not assessed
	Rijkenberg *et al.* (2015), the Netherlands	Descriptive design, convenience sample in ICU (n = 48 medical, n = 20 surgical), 1088 assessments by trained nurse raters	Discriminant validity: overall change in median score from rest (3·0, range 3·0–3·0; p = <0·001) to painful procedure (5·0, range 4·0–6·0; p = <0·001). Increased median BPS scores between rest (3·0, range 3·0–3·0) and nonpainful stimuli (3·0, range 3·0–4·0; $p = 0{\cdot}002$). Increase in median BPS score in deeply sedated subgroup (group 1, n = 28, RASS −5 to −3), rest (3·0, range 3·0–3·5) to painful procedure (5·0, range 4·0–6·8; $p < 0{\cdot}001$); lightly sedated subgroup (group 2, n = 33, RASS −2 to 0), rest (3·0, range 3·0–3·5) to nonpainful stimuli (4·0, 3·0–4·5; $p = 0{\cdot}003$)	Internal consistency: Cronbach's $\alpha = 0{\cdot}70$ during painful procedure. Inter-rater reliability: ICC = 0·74 (95%, CI 0·68–0·79; $p = 0{\cdot}001$).	Not assessed	Not assessed
Critical-Care Pain		Descriptive three-phase study,	Construct validity: mean pain intensity scores were higher during	Inter-rater reliability: $K_w = 0{\cdot}52–0{\cdot}88$	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
Observation Tool	Gélinas *et al.* (2006), Canada	convenience sample (n = 105 ICU patients), 315 assessments	positioning (2·01) than during rest (1·71) and recovery (1·40); patients' self-reporting correlated with pain intensity scores at extubation (rest, σ = 0·49; procedure σ = 0·59; recovery, σ = 0·40; $p < 0·001$). Discriminant validity: change in score at rest preprocedure, unconscious intubated (n = 33, t = −9·07; $p < 0·001$), conscious intubated (n = 99, t = −12·07; $p < 0·001$) and conscious extubated (n = 105, t = −15·96; $p < 0·001$). Specificity and sensitivity: painful procedures 86% and 78%, respectively; at rest, 83% and 83%, respectively			
	Gélinas and Johnston (2007), Canada	Descriptive design, convenience sample (n = 55 ICU patients), by trained 51 nurse raters.	Concurrent validity: patients' self-reporting of pain, high correlation (r = 0·71; $p \leq 0·05$). Sensitivity and specificity: CPOT cut-off score of >3 yielded a sensitivity of 66·7% and a specificity of 83·3%. Positive predictive value of 85·7% at CPOT score >3. Discriminant validity: MANOVA testing, CPOT score and physiologic indicators compared to painful (turning) and nonpainful stimuli (NIPB monitoring), significant difference found for the CPOT scores and physiologic indicators found for painful stimuli (F = 59·35–8·34; $p < 0·001$)	Inter-rater reliability: ICC = 0·80–0·93	Not assessed	Not assessed
	Gélinas *et al.* (2009), Canada	Descriptive repeated-measures study, convenience sample (n = 105, cardiovascular ICU patients), 315 assessments	Sensitivity and specificity: CPOT cut-off >2 during painful exposure; during painful stimuli, sensitivity 86%, specificity 78%; positive LR of 3·87 (1·63–9·23), negative LR of 0·18 (0·09–0·33); during nonpainful stimulation, specificity 83% and 97%, lower sensitivity 47% and 63%.	Not assessed	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
	Gélinas (2010), Canada	Descriptive design, convenience sample (n = 33 returned evolution forms)	Not assessed	Not assessed	Feasibility: 78·8–100%. Utility: 54·5–72·7%	Not assessed
	Marmo and Fowler (2010), USA	Descriptive repeated-measures study, convenience sample (n = 25 ICU patients), 300 assessments by four trained nurse raters	Not assessed	Internal consistency: Cronbach's α = 0·89. Inter-rater agreement: 56–100%	Not assessed	Not assessed
	Gélinas *et al.* (2011), Canada	Descriptive pre- and postimplementation study, convenience sample (pre, n = 141; 3 months, n = 65; 12 months, n = 44)	Not assessed	Inter-rater reliability: three video vignettes of patient at rest (95–100%) both pre- (n = 60) and postimplementation (n = 29); three video vignettes of patient being turned (painful stimuli) lower (73·3–91·7%) agreement pre-implementation compared to postimplementation (86·1–100%)	Document analysis at 3 and 12 months postimplementation: Three- to fourfold increase in reporting of pain assessments sustained at 3- and 12-month period (U = 385·00; p = 0·34). Decrease administration of sedative (propofol) boluses (H = 10·29; p < 0·05) postimplementation; analgesia boluses decreased but less significantly (H = 5·84; p = <0·10). Increase in pain reassessment following administration of analgesia sustained over 3–12 months postimplementation (U = 298·00; p = 0·97). Decrease CPOT (≥2) score postimplementation, over 75% of pharmacological interventions documented as effective at 3 and 12 months (U = 106·00; p = 0·43), with less behavioural indicators documented implementation of CPOT	Not assessed
	Damström *et al.* (2011), Sweden	Descriptive study, convenience sample (n = 40 ICU patients), 240 assessments by two trained nurse raters	Construct validity: change in CPOT score and correlation with MAP (ρ = 0·32–0·45; p < 0·05). Discriminant validity: multivariate analysis revealed significant effects for both Procedure (Wilk's λ = 0·75, $F_{4,34}$ = 2·91, p < 0·05) and	Internal consistency: nonpainful stimuli, Cronbach's α = 0·56–72, ICC = 0·73–92; painful stimuli, Cronbach's α = 0·31–81, ICC = 0·80–0·91.	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
			assessment occasions (Wilk's $\lambda = 0{\cdot}60$, $F_{8,34} = 5{\cdot}14$, $p < 0{\cdot}05$), as well as an interaction (procedure x occasion) (Wilk's $\lambda = 0{\cdot}60$, $F_{8,34} = 5{\cdot}18$, $p < 0{\cdot}05$)	Inter-rater reliability: overall ICC = 0·84 (0·72–0·92)		
	Paulson-Conger *et al.* (2011), USA	Descriptive, prospective, comparative study, convenience sample (n = 100) in ICU (n = 13 cardiac ICU, n = 27 medical ICU, n = 26 surgical ICU, and n = 24 neurological ICU), 200 assessments (n = 100 CPOT, n = 100 PAINAD) by four nurse raters	Discriminant validity: comparison with PAINAD scores high (0·86; $p < 0{\cdot}001$); Bland–Altman plot, mean bias was 0·10 (SD 1·22)	Internal consistence: Cronbach's $\alpha = 0{\cdot}76$	Not assessed	Not assessed
	Vazquez *et al.* (2011), Spain	Descriptive study, convenience sample (n = 96 ICU patients), 330 assessments, first 88 observation conducted simultaneously by two researchers	Criterion validity: change in mean score, preprocedure 0·27; during procedure 1·93 and postprocedure 0·10 (p = <0·05). Construct validity: comparison between mean scores and domains, before–during procedure (facial expression 0·10–0·80, body movements 0·04–0·43, muscle tension 0·08–0·34, compliance with ventilator 0·04–0·37; $p < 0{\cdot}01$); during–after (facial expression 0·80 –0·05, body movements 0·43–0·01, muscle tension 0·34–0·04, compliance with ventilator 0·37–0·01; $p < 0{\cdot}01$); and before–after (facial expression 0·10–0·05, muscle tension 0·08–0·04, compliance with ventilator 0·27–0·10; $p < 0{\cdot}01$)	Inter-rater reliability: $K = 0{\cdot}79$–$1{\cdot}0$	Not assessed	Not assessed
	Linde *et al.* (2013), USA	Descriptive study, convenience sample (n = 30	Construct validity: dressing change and turning (increase, +2·80; 95% CI 1·84–3·75; $p < 0{\cdot}001$)	Inter-rater reliability: overall, K = 0·87; mean rater scores, dressing change (increase, + 0·25; 95% CI −0·07 to 0·57;	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
		cardiothoracic ICU patients)		$p = 0{\cdot}12$); during turning (increase, +3·04; 95% CI 2·11–3·98; $p < 0{\cdot}001$)		
	Topolovec-Vranic *et al.* (2013), Canada	Descriptive study, convenience sample ($n = 66$ ICU patients), by 23 nurse raters	Concurrent validity: patients' (32/34) self-report of pain at rest 37·5% and postprocedure 43·8%; painful procedure 62·5%. Correlation between patient's (32/34) self-report and CPOT score $\sigma = 0{\cdot}435$; $p = <0{\cdot}001$ (two-tailed). Discriminant validity: repeated-measures ANOVA, turning, noncommunicative patients $F = 5{\cdot}81$; $p = 0{\cdot}019$; BP cuff inflation $F = 0{\cdot}10$; $p = 0{\cdot}756$	Inter-rater reliability: ICC 0·60–0·97	Ten completed questionnaires comparing CPOT to NVPS; feasibility (sufficiently very 49% vs. 47%, respectively), applicability (sufficiently very, 18% vs. 7%, respectively)	Not assessed
	Chanques *et al.* (2014), USA	Descriptive study, convenience sample ($n = 30$ ICU patients), 258 paired assessments, by 24 raters ($n = 20$ nurses, $n = 4$ observers)	Discriminant validation: CPOT score increases from baseline to procedure in line with BPS and NVPS scores and 10 min^{-1} after procedure ($p < 0{\cdot}001$), nil (median) score differences noted ($p > 0{\cdot}05$). Difference in pain scores between BPS, CPOT and NVPS not significant ($p > 0{\cdot}05$)	Internal consistency: Overall, Cronbach's $\alpha = 0{\cdot}81$, intubated patients Cronbach's $\alpha = 0{\cdot}82$, nonintubated Cronbach's $\alpha = 0{\cdot}81$. Inter-rater reliability: Total score $K_w = 0{\cdot}81$. Domains: face $K_w = 0{\cdot}81$; breathing domains, ventilation/vocalisation $K_w = 0{\cdot}71$; muscular domains, body movements $K_w = 0{\cdot}42$ and muscle tension $K_w = 0{\cdot}43$	Not assessed	Not assessed
	Rijkenberg *et al.* (2015), the Netherlands	Descriptive study, convenience sample ($n = 48$ medical ICU; $n = 20$ surgical ICU patients), 1088 assessments by nurse raters	Discriminant validity: overall change in median score from rest (0·0, range 0·0–0·0) to painful procedure (2·0, range 0·0–3·0; $p = <0{\cdot}001$). Unchanged median CPOT scores between rest (0·0, range 0·0–0·0) and nonpainful stimuli (0·0, range 0·0–2·0; $p = 0{\cdot}002$). Increase in median CPOT score in deeply sedated subgroup (group 1, $n = 28$, RASS −5 to −3), rest (0·0, range 0·0–0·0) to painful procedure (2·0, range 0·3–4·0; $p = <0{\cdot}001$); median CPOT score stable in lightly sedated subgroup (group 2, $n = 33$, RASS −2 to 0) rest (0·0, range 0·0–	Internal consistency: Cronbach's $\alpha = 0{\cdot}71$ during painful procedure. Inter-rater reliability: ICC = 0·74 (95%, CI 0·68–0·79; $p = 0{\cdot}001$)	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
			0·0) to nonpainful stimuli (0·0, range 0·0–1·5; p = 0·006)			
	Kanji *et al.* (2016), Canada	Prospective cohort study, convenience sample (n = 40 ICU patient with delirium)	Discriminant validity: change in mean CPOT scores between baseline and painful procedures (3·13 ± 1·56; p < 0·001; Cohen D, 2·0)	Internal consistency: Cronbach's α = 0·778 Inter-rater reliability: k > 0·60 with high correlation between rater total scores (r = 0·957) and subjective opinion compared to physiologic variables (80·5% vs. 67·5%, respectively)		
Faces, Legs, Activity, Cry and Consolability Scale (FLACC)	Odhner *et al.* (2003), USA	Descriptive study, convenience sample (n = 59 ICU patients), 100 paired assessments (200 individual)	Convergent validity: correlation between FLACC and NVPS items; face (r = 0·78), FLACC legs and NVPS activity item (r = 0·72), FLACC total score and NVPS face (r = 0·71), activity (r = 0·71) and guarding items (r = 0·69; NVPS total score and FLACC consolability item (r = 0·75), legs (r = 0·73), activity (r = 0·71) and face components (r = 0·65); total FLACC and NVPS scores (r = 0·86)	Internal consistency: Cronbach's α = 0·84	FLACC cry and consolability often marked 'not applicable' (62 and 16 times, respectively)	Not assessed
	Voepel-Lewis *et al.* (2010), USA	Descriptive study, convenience sample (n = 29 ICU adult patients), 60 assessments	Construct validity: change in score postadministration of analgesia (mean, 5·27; SD 2·3 vs. mean 0·52; SD 1·1; p < 0·001). Factor analysis: first factor accounted for 68·9% of variance in FLACC scores; four items contributed to this component: face (0·86), legs (0·94), activity (0·90) and consolability (0·95)	Internal consistence: overall Cronbach'sα = 0·882; each category highly correlated with the others (ρ = 0·69–0·92; p < 0·001) expect the cry category (ρ = 0·18–0·36). Cronbach's α improved to 0·934 when cry category removed. Inter-rater reliability: Adult sample, face, K = 0·93; Legs, K = 0·97; Activity, K = 0·93; Cry, K = 0·72; Consolability, K = 0·96; total score, K = 0·98	Not assessed	Not assessed
Pain Assessment in	Paulson-Conger *et al.* (2011), USA	Descriptive, prospective, comparative study,	Convergent validity: comparison with CPOT scores high (0·86;	Internal consistency: Cronbach's α=0·80	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
Advanced Dementia		convenience sample (n = 13 cardiac ICU, n = 27 medical ICU, n = 26 surgical ICU, and n = 24 neurological ICU), 200 assessments (n = 100 CPOT, n = 100 PAINAD) by four nurse raters	$p < 0{\cdot}001$); Bland–Altman plot, mean bias was 0·10 (SD 1·22)			
Nonverbal Pain Scale (NVPS)	Odhner *et al.* (2003), USA	Descriptive study, convenience sample (n = 59 ICU patients), 100 paired assessments (200 individual)	Convergent validity: correlation between FLACC and NVPS items; face (r = 0·78), FLACC legs and NVPS activity item (r = 0·72), FLACC total score and NVPS face (r = 0·71), activity (r = 0·71) and guarding items (r = 0·69; NVPS total score and FLACC consolability item (r = 0·75), legs (r = 0·73), activity (r = 0·71) and face components (r = 0·65); total FLACC and NVPS scores (r = 0·86).	Internal consistency: Cronbach's α = 0·78	Not assessed	Not assessed
	Ahlers *et al.* (2008), the Netherlands	Descriptive study, convenience sample (n = 113 ICU patients), 371 assessments	Not assessed	Inter-rater reliability: K = 0·71	Not assessed	Not assessed
	Kabes *et al.* (2009), USA	Descriptive study, comparative study, comparative convenience sample (n = 64 ICU patients), 121 independent assessments by nine trained nurse raters	Construct validity: change in score between rating before, during and after painful stimulus F = 135·86; $p < 0{\cdot}001$.	Internal consistency: (original NVPS) before stimulus Cronbach's α = 0·36, during stimulus Cronbach's α = 0·62 and after stimulus Cronbach's α = 0·62. Inter-rater reliability: Above 90% agreement between two raters achieved >94·7% on 72 observations	Not assessed	Not assessed
	Marmo and Fowler (2010), USA	Descriptive study, repeated-measures study, convenience sample (n = 25 ICU patients), 300	Not assessed	Internal consistency: Cronbach's α = 0·89. Inter-rater agreement: suctioning 78% (58/75 assessments), turning 79%	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Table 3 (continued)

Instrument	Study, country	Design, setting, sample	Validity	Reliability	Applicability	Responsiveness
		assessments by four trained nurse raters		(59/75 assessments); highest disagreement by scale item, on suctioning, face (9%, 13 of 75), on turning, guarding/ muscle tension (6%, 9 of 75)		
	Juarez *et al.* (2010), USA	Descriptive study, convenience sample (n = 124 surgical ICU; n = 115 medical ICU), 968 assessments	Construct validity: change in score (rest, 3·73 vs. procedures, 5·41; p = <0·001); change from baseline, $Z = -11{\cdot}125$; $p < 0{\cdot}001$). Convergent validity: Correlation with NVPS (rest $\rho = 0{\cdot}69$, procedure, $\rho = 0{\cdot}77$; $p < 0{\cdot}001$)	Internal consistency: overall, Cronbach's $\alpha = 0{\cdot}75$	Not assessed	Not assessed
	Topolovec-Vranic *et al.* (2010), Canada	Descriptive study, convenience sample (n = 66 ICU patients), by 23 nurse raters	Concurrent validation: patients' (32/34) self-report of pain at rest 37·5% and postprocedure 43·8%; painful procedure 62·5%. Correlation between patient's (32/34) self-report and NVPS score $\sigma = 0{\cdot}313$; $p < 0{\cdot}001$ (two-tailed). Discriminant validity: repeated-measures ANOVA, turning, noncommunicative patients $F = 5{\cdot}32$, $p = 0{\cdot}025$; BP cuff inflation $F = 0{\cdot}10$, $p = 0{\cdot}756$	Inter-rater reliability: ICC 0·34–0·97	Ten completed questionnaires, comparing CPOT to NVPS; feasibility (sufficiently very 49% vs. 47%, respectively), applicability (sufficiently very, 18% vs. 7%, respectively)	Not assessed
	Chanques *et al.* (2014), USA	Descriptive study, convenience sample (n = 30 ICU patients), 258 paired assessments, by 24 raters (n = 20 nurses, n = 4 observers)	Discriminant validation: NVPS score increases from baseline to procedure in line with BPS and CPOT scores and 10 min^{-1} after procedure ($p < 0{\cdot}001$), nil (median) score differences noted ($p > 0{\cdot}05$). Difference in pain scores between BPS, CPOT and NVPS not significant ($p > 0{\cdot}05$)	Internal consistency: Overall, Cronbach's $\alpha = 0{\cdot}81$, intubated patients Cronbach's $\alpha = 0{\cdot}82$, nonintubated Cronbach's $\alpha = 0{\cdot}81$. Inter-rater reliability: Total score $K_w = 0{\cdot}71$. Domains: face $K_w = 0{\cdot}70$; respiration $K_w = 0{\cdot}54$; activity $K_w = 0{\cdot}52$; guarding $K_w = 0{\cdot}32$; physiology I $K_w = 0{\cdot}45$; and physiology II $K_w = 0{\cdot}02$	Not assessed	Not assessed

© 2016 John Wiley & Sons Ltd

Quality of the studies

The quality of the studies included in the systematic review was fair to high (median 11, range 8–12). Convenience sampling methods were judged to be appropriate, with suitable inclusion and exclusion criteria stated. The majority ($n = 23$; 88%) of studies provided a detailed description of the order and type of patient stimuli, including duration of assessment. Assessments were largely ($n = 20$; 77%) undertaken by trained independent assessors. Seven studies were able to obtain patient self-report of pain to assess accuracy of outcome measured. Common limitations concerned the identification and analysis of confounding factors and sample size.

In the review, potential confounders considered pertinent were patient depth of sedation and degree of delirium. Ten of the studies included in the review measured depth of sedation at the time of assessing pain response (Payen *et al.* 2001, Aïssaoui *et al.* 2005, Young *et al.* 2006, Gélinas & Johnston 2007, Ahlers *et al.* 2008, Juarez *et al.* 2010, Chen *et al.* 2011, Damström *et al.* 2011, Morete *et al.* 2014, Rijkenberg *et al.* 2015). While no study examined whether depth of sedation potentially confounded the level of pain measured, one study reported a moderate negative correlation ($r = 0{\cdot}432$; $p < 0{\cdot}001$) between depth of sedation and level of pain detected using the BPS (Aïssaoui *et al.* 2005). Only one study (Rijkenberg *et al.* 2015) assessed for patient delirium, but did not examine whether the degree of delirium potentially confounded the level of pain measured (Table 4).

Evidence of validity, reliability, applicability and responsiveness

Behavioural Pain Scale

Initially developed in France, the BPS has been translated into English (Payen *et al.* 2001), Chinese (Chen *et al.* 2011), Finnish (Pudas-Tähkä *et al.* 2014), as well as Brazilian Portuguese (Morete *et al.* 2014). Validity and reliability has been evaluated in critically ill intubated patients across ten studies, in different intensive care settings comprising a total of 940 mechanically ventilated adult patients (Payen *et al.* 2001, Aïssaoui *et al.* 2005, Young *et al.* 2006, Ahlers *et al.* 2008, 2010, Juarez *et al.* 2010, Chen *et al.* 2011, Chanques *et al.* 2014, Morete *et al.* 2014, Rijkenberg *et al.* 2015).

Of these ten studies, six examined validity (Payen *et al.* 2001, Aïssaoui *et al.* 2005, Young *et al.* 2006, Ahlers *et al.* 2010, Juarez *et al.* 2010, Chen *et al.* 2011). For construct validity, two studies (Payen *et al.* 2001, Ahlers *et al.* 2010) reported change in mean total score from baseline to painful stimuli (e.g. turning, endotracheal suctioning). Statistically significant (3·0 vs. 6·8; $p < 0{\cdot}05$) changes in mean scores from baseline to painful stimuli were reported (Aïssaoui *et al.* 2005, Rijkenberg *et al.* 2015) but not between baseline and nonpainful stimuli (3·1 vs. 3·5; $p > 0{\cdot}05$) (Payen *et al.* 2001, Ahlers *et al.* 2010). For convergent validity, scores compared favourably to NVPS (rest, $\rho = 0{\cdot}69$, procedure, $\rho = 0{\cdot}77$; $p < 0{\cdot}001$) in one study (Juarez *et al.* 2010). Instrument domains converged on a single pain expression factor during principal component factor analysis, explaining 55–65% of total variance (Payen *et al.* 2001, Aïssaoui *et al.* 2005). Concurrent validity was established by comparing BPS scores to patients' ($n = 30$) self-reports of pain intensity (Ahlers *et al.* 2010) and showed a positive and statistically significant correlation ($\rho = 0{\cdot}67$; $p = <0{\cdot}001$).

Reliability was evaluated across nine studies using a variety of approaches (Payen *et al.* 2001, Aïssaoui *et al.* 2005, Young *et al.* 2006, Ahlers *et al.* 2008, 2010, Juarez *et al.* 2010, Chanques *et al.* 2014, Morete *et al.* 2014, Rijkenberg *et al.* 2015). Internal consistency varied, with Cronbach's α coefficients ranging from 0·60–0·83 (Juarez *et al.* 2010) and interclass coefficients of 0·50–1·00 reported (Aïssaoui *et al.* 2005, Young *et al.* 2006, Chen *et al.* 2011). Total score demonstrated fair-to-moderate inter-rater reliability ($\kappa = 0{\cdot}67$–$0{\cdot}83$, $\kappa_w = 0{\cdot}81$) (Ahlers *et al.* 2008, 2010, Chanques *et al.* 2014) and moderate-to-high interclass agreement (ICC = 0·74–0·94) (Rijkenberg *et al.* 2015). Assessment of the upper limb domain demonstrated lowest agreement ($K = 0{\cdot}61$–$0{\cdot}72$) (Ahlers *et al.* 2010, Chanques *et al.* 2014). Comparisons of BPS total score to CPOT (Chanques *et al.* 2014) and NVPS (Juarez *et al.* 2010, Chanques *et al.* 2014) demonstrated no significant variation between pain intensity scores from rest to painful stimuli ($p > 0{\cdot}05$). In a mixed intensive care environment (Juarez *et al.* 2010), inter-rater accuracy when comparing the BPS and NVPS was lower (ICC = 0·46–0·58) for medical compared to surgical patients (ICC 0·68–0·73) and was highest for nonpainful procedures (eye care, 64% to 73%), compared to painful stimuli (turning, 36–46%) (Young *et al.* 2006).

Feasibility and utility testing found that the majority of intensive care clinicians ($n = 28$, 86%) were satisfied with ease of use (Payen *et al.* 2001). Sensitivity and specificity was examined in one study evaluating a Chinese language version and found that for a score of more than 6·5, the area under the curve of 75·9% indicated discrimination, with 52·4% sensitivity and 87·5% specificity (Chen *et al.*

© 2016 John Wiley & Sons Ltd

Table 4 Summary of Critical Appraisal Skills Program results

Reference	Clear aim	Acceptable sampling	Exposure accurately measured	Outcome accurately measured	Confounding factor(s) identified and accounted for	Follow-up sufficiently complete	Results clearly described	Reporting of probability values	Results believable	Results applicable to local population	Results fit with other evidence	Implications for practice	Total score (%)
Payen *et al.* (2001)	+	+	+	+	+[S]	+	+	+	+	+	+	+	12 (100)
Odhner *et al.* (2003)	+	+	+	+	–	–	+	+	+	+	–/+	+	9 (75·0)
Aïssaoui *et al.* (2005)	+	+	+	+	+[S]	+	+	+	+	+	+	+	12 (100)
Young *et al.* (2006)	+	+	+	–	+[S]	+	+	+	+	+	+	+	11 (92)
Gélinas *et al.* (2006)	+	+	+	+[P]	–	+	+	+	+	+	+	+	11 (92)
Gélinas and Johnston (2007)	+	+	+	+[P]	+[S]	+	+	+	+	+	+	+	11 (92)
Ahlers *et al.* (2008)	+	+	+	+	+[S]	+	+	+	+	+	+	+	12 (100)
Gélinas *et al.* (2009)	+	+	+	+[P]	–	+	+	+	+	+	+	+	11 (92)
Kabes *et al.* (2009)	+	+	+	–	–	+	+	+	+	+	+	+	10 (83)
Ahlers *et al.* (2010)	+	+	+	+	+/–	+	+	+	+	+	+	+	11 (92)
Juarez *et al.* (2010)	+	+	+	+	+[S]	+/–	+	+	+	+	+	+	11 (92)
Gélinas (2010)	+	+	+	+	+	+	+	+	+	+	+	+	11 (100)
Marmo and Fowler (2010)	+	+	+	+	–	+	+	+	+	+	+	+	11 (92)
Voepel-Lewis *et al.* (2010)	+	+	+	+	–	+	+	+	+	+	+	+	11 (92)
Topolovec-Vranic *et al.* (2010)	+	+	+	+[P]	–	+	+	+	+	+	+/–	+	9 (75)
Chen *et al.* (2011)	+	+	+	+[P]	+[S]	+/–	+	+	+	+	+	+	11 (92)
Gélinas *et al.* (2011)	+	+	+	+/–	–	+	+	+	+/–	+	+	+	9 (75)
Damström *et al.* (2011)	+	+	+	+[P]	+[S]	+	+	+	+	+	+	+	11 (92)
Paulson-Conger *et al.* (2011)	+	+	+	–	–	+/–	+	+	+	+	+	+	9 (75)
Vazquez *et al.* (2011)	+	+	+	+/–	–	+	+	+	+	+	+	+	10 (83)
Linde *et al.* (2013)	+	+	+	+	–	+	+	+	+	+	+	+	11 (92)
Topolovec-Vranic *et al.* (2013)	+	+	+/–	+/–[P]	–	+	+	+	+	+	+	+	10 (83)
Chanques *et al.* (2014)	+	+	+	+[R]	–	+	+	+	+	+	+	+	11 (92)
Morete *et al.* (2014)	+	+	+/–	+	+[S]	–	+/–	+	+	+	+	+	8 (67)
Rijkenberg *et al.* (2015)	+	+	+/–	+	+[S, D]	+	+	+	+	+	+	+	11 (92)
Kanji *et al.* (2016)	+	+	+	+	–	+	+	+	+	+	+	+	10 (83)

D, delirium screen; P, patient self-report obtained; R, randomised order of assessment; and S, sedation level.

© 2016 John Wiley & Sons Ltd

2011). Responsiveness was examined in one study based on 30 intensive care patients, demonstrating a large effect size (2·2–3·4) for three subscale scores and BPS scores in quantifying and detecting painful procedures (Aïssaoui *et al.* 2005).

While the BPS has demonstrated acceptable validity and reliability properties across multiple studies, inherent limitations remain regarding the operational definition of some items (e.g. movements of upper arms, compliance with the ventilator) that may be interpreted differently between clinical staff. Only fair-to-moderate internal consistency (Cronbach's $\alpha = 0{\cdot}60$–$0{\cdot}77$) (Juarez *et al.* 2010) potentially suggests variation in user interpretation of the items. A further limitation concerns the number of intensive care studies (Payen *et al.* 2001, Juarez *et al.* 2010, Chen *et al.* 2011) that used repeated observations within subjects to calculate coefficients, potentially leading to higher results being reported (Abdi 2007, Streiner *et al.* 2015).

Critical-Care Pain Observation Tool

Originally developed in French (Gélinas *et al.* 2006) with forward and back translations into English (Gélinas & Johnston 2007), Spanish (Vazquez *et al.* 2011), Swedish (Damström *et al.* 2011) and Finnish (Pudas-Tähkä *et al.* 2014), a total of 14 studies examined the CPOT in different intensive care settings (i.e. surgical, medical, trauma and neurological) for a total of 850 adult patients (Gélinas *et al.* 2006, 2009, Gélinas & Johnston 2007, Gélinas 2010, Marmo & Fowler 2010, Damström *et al.* 2011, Gélinas *et al.* 2011, Paulson-Conger *et al.* 2011, Vazquez *et al.* 2011, Linde *et al.* 2013, Topolovec-Vranic *et al.* 2013, Chanques *et al.* 2014, Rijkenberg *et al.* 2015, Kanji *et al.* 2016).

Various aspects of validity were examined across ten studies (Gélinas *et al.* 2006, Gélinas & Johnston 2007, Damström *et al.* 2011, Paulson-Conger *et al.* 2011, Vazquez *et al.* 2011, Linde *et al.* 2013, Topolovec-Vranic *et al.* 2013, Chanques *et al.* 2014, Rijkenberg *et al.* 2015, Kanji *et al.* 2016). Content validity was established through expert review ($n = 4$ physicians, $n = 13$ intensive care nurses) via questionnaire, rating the relevance of items using a four-point Likert as high (0·88–1·0) (Gélinas *et al.* 2006). Concurrent validity was also supported, with significant moderate-to-high correlations ($r = 0{\cdot}71$; $p < 0{\cdot}05$) between CPOT scores and patient self-report ($n = 55$) of pain intensity at rest (Gélinas & Johnston 2007). Discriminant validity was assessed in two studies (Damström *et al.* 2011, Kanji *et al.* 2016), demonstrating statistically significant change in mean scores from baseline compared to painful procedure (mean difference, $3{\cdot}13 \pm 1{\cdot}56$; $p < 0{\cdot}001$; Cohen D, 2·0), and increases in scores of two to three points from rest (Wilk's $\lambda = 0{\cdot}75$, $F_{4,34} = 2{\cdot}91$, $p = <0{\cdot}05$) to painful procedures ($\lambda = 0{\cdot}60$, $F_{8,34} = 5{\cdot}14$, $p = <0{\cdot}05$) and interaction ($\lambda = 0{\cdot}60$ $F_{8,34} = 5{\cdot}18$, $p = <0{\cdot}05$). Convergent validity was examined in three studies (Paulson-Conger *et al.* 2011, Chanques *et al.* 2014, Rijkenberg *et al.* 2015) with statistically significant increase between scores from rest to application of painful stimuli noted on comparison to BPS (Chanques *et al.* 2014, Rijkenberg *et al.* 2015) and PAINAD (Paulson-Conger *et al.* 2011) scores.

Reliability was examined in 12 intensive care studies (Gélinas *et al.* 2006, Gélinas & Johnston 2007, Marmo & Fowler 2010, Damström *et al.* 2011, Gélinas *et al.* 2011, Paulson-Conger *et al.* 2011, Vazquez *et al.* 2011, Linde *et al.* 2013, Topolovec-Vranic *et al.* 2013, Chanques *et al.* 2014, Rijkenberg *et al.* 2015, Kanji *et al.* 2016). Internal consistency varied with Cronbach's α coefficients ranging from 0·31–0·81 (Damström *et al.* 2011, Chanques *et al.* 2014). Moderate-to-high measures of inter-rater reliability from two or more raters were found across four translated versions: French, $\kappa_w = 0{\cdot}52$–$0{\cdot}88$ (Gélinas *et al.* 2006); English, ICC = 0·74–0·93, $r = 0{\cdot}957$, $\kappa > 0{\cdot}60$ (Gélinas & Johnston 2007, Paulson-Conger *et al.* 2011, Chanques *et al.* 2014, Rijkenberg *et al.* 2015, Kanji *et al.* 2016); Spanish, $\kappa = 0{\cdot}79$–$1{\cdot}0$ (Vazquez *et al.* 2011); and Swedish, ICC = 0·72–92 (Damström *et al.* 2011). Based on 258 paired assessments of critically ill patients ($n = 30$), inter-rater reliability of the CPOT, whilst high (0·86), was found to be lower during patient repositioning and turning (painful) procedures when compared to the BPS (0·90) and NVPS (0·92) (Chanques *et al.* 2014). Sensitivity of 67–86% and specificity of 78–83% were noted for a cut-off score between 2–3 to detect the presence of significant pain (Gélinas & Johnston 2007, Gélinas *et al.* 2009).

Feasibility and clinical utility were assessed in two studies (Gélinas 2010, Chanques *et al.* 2014) of the CPOT. A survey involving intensive care nurses ($n = 33$) and respondents ($n = 24$; 73%) indicated that they would recommend the use of the CPOT as helpful for nursing practice; however, not all nursers ($n = 28$; 54%) reported that the CPOT positively influenced their practice in assessing pain (Gélinas 2010). Further, while the directives regarding the CPOT's use were clear and simple ($n = 33$; 100%), this did not translate into being quick to use ($n = 26$; 79%). On comparing CPOT to the BPS and NVPS, critically care nurses ($n = 20$) rated the BPS significantly higher (median 8, range 7–10, $p = 0{\cdot}02$) compared to the CPOT (median 8, range 5–8), but least preferred pain measure ($n = 5$; 24%) (Chanques *et al.* 2014).

© 2016 John Wiley & Sons Ltd

With extensive testing of CPOT, some other limitations are noted. Testing in different critically ill intensive care patient groups while broad often involved small sample sizes limiting representativeness. While this is one of only two pain measures of pain developed for use in both verbal and nonverbal patients, the vocalisation item has only been tested with postoperative cardiac surgery patients (n = 30) who were able to self-report following extubation. The instrument also requires further testing with different patient groups unable to self-report, such as those with delirium, dementia or intellectual deficits. Further, as with all instruments reliant upon facial expression as an indicator that pain is present, further testing should include brain-injured patients, as many present with different facial expressions compared to those exposed to painful procedures (Roulin & Ramelet 2012).

Faces, Legs, Activity, Cry and Consolability Scale

While initially designed to assess for pain in children with cognitive impairment (Merkel *et al.* 1997), two studies have explored the use of the FLACC instrument in a total of 87 critically ill intubated adults (Odhner *et al.* 2003, Voepel-Lewis *et al.* 2010). Construct validity was demonstrated through significant decreases in scores from painful stimuli to rest and following administration of analgesia (mean, 5·3; SD 2·3 vs. mean, 0·52; SD 1·1; p = <0·001, respectively) (Voepel-Lewis *et al.* 2010). Concurrent validity was demonstrated with high correlation between total FLACC and NVPS scores (r = 0·71–0·86) and FLACC domain and NVPS total scores (r = 0·65–0·78; p < 0·01) (Odhner *et al.* 2003). An explanatory factor analysis identified contributions to score variance: the lowest was for the cry domain (68·5%) compared to face (0·86), legs (0·94), activity (0·90) and consolability (0·95). High internal consistency was demonstrated (Cronbach's α = 0·84–88), which increased to 0·934 when the cry domain was removed (Voepel-Lewis *et al.* 2010). Inter-rater reliability was high (total score K = 0·98).

No studies were identified that explored responsiveness or feasibility and utility with critically ill adult patients. In addition to further research involving larger samples across a range of diverse clinical settings (i.e. surgical, medical and neurotrauma), criterion validity could be improved by comparing scores against patient self-report (International Association for the Study of Pain, 2006, National Health and Medical Research Council, 2011), instead of another pain measure.

Pain Assessment in Advanced Dementia

Originally developed for use in patients whose dementia is so advanced that they cannot verbally communicate, the PAINAD instrument has recently been compared with the CPOT. In one study (Paulson-Conger *et al.* 2011), the PAINAD instrument was evaluated for potential use in critically ill intubated adult patients (n = 100) drawn from four specialist ICUs (cardiac, n = 13; medical, n = 27; surgical, n = 26; and neurological, n = 34). Internal reliability was higher than the CPOT (ICC 0·80 vs. 0·76), with high and significant correlations between the two instruments (0·86; p = <0·001). Further assessment of agreement using Bland–Altman plots demonstrated that the two measures were interchangeable (variance less than ±2 SD) (Mantha *et al.* 2000).

Limitations were however noted. While internal consistency and convergent validity was examined in the above study, inter-rater reliability, internal reliability, discriminant validity and criterion validity have not been re-examined in the intensive care setting. While this has been described elsewhere in aged care (Warden *et al.* 2003, Leong *et al.* 2006, DeWaters *et al.* 2008), the degree to which this can be extrapolated to critical care areas is unknown. Both PAINAD and CPOT have been designed for different populations, clinical conditions and pathologies (Streiner *et al.* 2015). With the increasing number of older patients with cognitive impairment presenting to ED potentially requiring critical care, intubation and analgosedation (Hirschman *et al.* 2011, Australian Commission on Safety and Quality in Health Care, 2013), further research and psychometric evaluation is therefore warranted regarding PAINAD in critical care settings with intubated patients.

Nonverbal Pain Scale

The NVPS domains, in its first iteration, were based on the FLACC and included face (expression/grimacing), activity, guarding, physiology I and physiology II (Odhner *et al.* 2003). Domains of the revised instrument are listed in Table 2. Seven studies (Odhner *et al.* 2003, Ahlers *et al.* 2008, Kabes *et al.* 2009, Juarez *et al.* 2010, Marmo & Fowler 2010, Topolovec-Vranic *et al.* 2010, Chanques *et al.* 2014) examined validity and reliability in critically ill patients for a total of 596 patients. Five studies examined validity (Odhner *et al.* 2003, Kabes *et al.* 2009, Juarez *et al.* 2010, Topolovec-Vranic *et al.* 2010, Chanques *et al.* 2014). As noted earlier, convergent validity between FLACC domains and NVPS total scores and FLACC total score and NVPS domains showed high correlations (r = 0·71–0·86) (Odhner *et al.* 2003). Discriminant validity was also supported (Kabes *et al.* 2009, Topolovec-Vranic *et al.* 2013), reflecting detection of changes in the presence of pain (z = 145·05, p < 0·001; F = 5·32, p = 0·025). Comparison of patient self-reports to NVPS scores (Topolovec-Vranic

© 2016 John Wiley & Sons Ltd

et al. 2013) demonstrated a moderate statistically correlation ($\rho = 0{\cdot}313$; $p = <0{\cdot}001$).

All seven studies also examined reliability. Internal consistency was moderate to high, with an overall Cronbach's α of 0·75, and 0·80 in surgical patients (Juarez *et al.* 2010). High inter-rater agreement (90·8%) was noted (Topolovec-Vranic *et al.* 2010) with a kappa of 0·71–0·81 reported (Ahlers *et al.* 2008, Chanques *et al.* 2014) although inter-class coefficients were wide (0·34–0·92; $p = <0{\cdot}001$). Feasibility was evaluated in one small ($n = 10$) study (Topolovec-Vranic *et al.* 2010), which found moderate-to-high feasibility (78–96%), but low (overall, 23%) acceptability when compared to CPOT.

Current published literature examining the NVPS pain has several limitations. Behavioural descriptors such as smiling or lying in a normal position cannot be solely equated with a nonpainful state. Further, measurement of the physiologic indicators pupil dilatation and perspiration was not defined or standardised, and there is no rationale for the selection of vital sign parameters. The study by Kabes *et al.* (2009) failed to report post hoc test data and statistical significance for the respiratory item. In the study by Topolovec-Vranic *et al.* (2010), there were three main limitations: 1) while staff confidence increased postimplementation, 38·5% fewer staff ($n = 20$) completed the postimplementation survey; 2) one of the researchers was also the Clinical Manager for the study site; and 3) there was potential selection bias as patient surveys were directed more towards patients who were able to complete the interview in a short space of time. To date, no study has been published examining the responsiveness of the NVPS in critical care adult populations.

Discussion

This systematic review identified five observational pain assessment instruments that were examined across a range of intensive care sites involving nonverbal critically ill patients. This review found that none of the observational pain assessment instruments have been tested in the ED setting. Yet, critically ill patients are increasingly presenting to the ED and often require intubation and mechanical ventilation, aggressive resuscitation and stabilisation (O'Connor *et al.* 2009). From 2008–2014, the number of patients presenting to Australian public hospital EDs has increased by nearly 1·1 m ($n = 1{,}078{,}659$; 20·2%), exceeding the rate of population growth (Australian Institute of Health and Welfare, 2008, 2015). In the same time period, the number of critically ill patients admitted to EDs has also increased by nearly 40% (Australasian Triage Scale 1 and 2, $n = 110{,}986$; 38·4%). While critically ill patients need to be transferred to an intensive care unit as swiftly as possible, this patient population often remains in the ED for prolonged periods of time (Richardson & Mountain 2009, Rose *et al.* 2012).

Pain is a common symptom experienced by the critically ill patient. In addition to pain associated with the initial injuries or conditions, critically ill patients are exposed to multiple other sources of pain during ED resuscitation and stabilisation (Weir & O'Neill 2008, Varndell *et al.* 2015b). As many as 79% of patients describe experiencing pain, most often of moderate-to-severe intensity whilst intubated and mechanically ventilated (Puntillo *et al.* 2001, Ma *et al.* 2010). Pain is a predominant stressor that can activate many pathophysiological and psychological mechanisms in critically ill patients, but assessment rates of pain in critical ill patients are suboptimal (Payen & Chanques 2012). This may be due to an inability of critically ill patients to communicate their pain due to sedation, cognitive impairment, paralysis, endotracheal intubation/mechanical ventilation or clinician workload. The infrequency of assessing, responding to and documenting pain is similar to that observed in intensive care settings, despite patients receiving analgesia (Hawksley 2000, Motov & Khan 2009, Thomas 2013, Varndell *et al.* 2015b). The positive association between the ability to assess and document patient pain and adequate management of pain has been previously documented (Payen *et al.* 2001, Mularski 2004). Conversely, inaccurate pain assessment and ensuing suboptimal treatment of pain in critically ill adults can affect all bodily systems with a plethora of physiologic and psychological consequences (Urden *et al.* 2010, Georgiou *et al.* 2015) involving short- and long-term clinical outcomes of critically ill patients (Dunwoody *et al.* 2008). This systematic review has identified a lack of evidence that EDs use observational pain assessment instruments for the critically ill patient. Consequently, there is an urgent need to examine the suitability and appropriateness of observation pain instruments for use in the nonverbal critically ill patient in ED.

This systematic review evaluated the suitability and psychometric properties of the instruments based on quality judgement criteria relating to validity, reliability and responsivity. A common weakness across the majority ($n = 18$; 69%) of studies (Payen *et al.* 2001, Odhner *et al.* 2003, Aïssaoui *et al.* 2005, Young *et al.* 2006, Gélinas & Johnston 2007, Gélinas *et al.* 2009, Kabes *et al.* 2009, Ahlers *et al.* 2010, Marmo & Fowler 2010, Topolovec-Vranic *et al.* 2010, 2013, Voepel-Lewis *et al.* 2010, Chen *et al.* 2011, Damström *et al.* 2011, Vazquez *et al.* 2011, Linde *et al.* 2013, Chanques *et al.* 2014, Rijkenberg *et al.*

© 2016 John Wiley & Sons Ltd

2015) was related to convenience sampling, small sample size and low number of item-to-subject (i.e. patient) ratio assessments, limiting generalisability of the instruments. Within the psychometric literature, a sample size of 100–200 is recommended (Bollen 1989, Anthoine *et al.* 2014). Further, an item-to-subject ratio of 1:10 is recommended for future testing of instruments (Schwab 1980). This systematic review identified five observational pain assessment instruments that were examined across a range of studies involving critically ill intubated patients. However, only three instruments (BPS, CPOT and PAINAD) were evaluated with sample sizes and item-to-subject ratios at these levels (Aïssaoui *et al.* 2005, Gélinas & Johnston 2007, Damström *et al.* 2011, Morete *et al.* 2014, Rijkenberg *et al.* 2015). Of these three instruments, the CPOT has had the most extensive testing for validity, reliability (Gélinas *et al.* 2006, 2009, Gélinas & Johnston 2007, Marmo & Fowler 2010, Damström *et al.* 2011, Gélinas *et al.* 2011, Paulson-Conger *et al.* 2011, Vazquez *et al.* 2011, Linde *et al.* 2013, Topolovec-Vranic *et al.* 2013, Chanques *et al.* 2014, Rijkenberg *et al.* 2015, Kanji *et al.* 2016) and feasibility (Gélinas 2010, Gélinas *et al.* 2011, Topolovec-Vranic *et al.* 2013), with high levels of sensitivity and specificity (Gélinas & Johnston 2007, Gélinas *et al.* 2009). The CPOT has also been tested against the BPS, NVPS and PAINAD (Paulson-Conger *et al.* 2011, Chanques *et al.* 2014) and the recommended gold standard, patient self-report (Odhner *et al.* 2003). Further, unlike the BPS and NVPS, the CPOT has been tested in both verbal and nonverbal patient populations, including patients with delirium, which potentially broadens the instruments use within the ED setting.

Strengths and limitations of the review

The strength of this study was the thorough and systematic search for relevant articles. To make this review more systematic and objective, the authors used standardised assessment frameworks to evaluate each study and on analysing the quality of psychometric testing. There are several limitations of this review. First, only one author (WV) extracted study data and made the initial critical appraisal of the selected instruments, although deposition and accuracy was verified by the other remaining authors. Second, descriptions of the development and testing of the observational pain assessment instruments vary in terms of quality and detail, which may limit generalisability of the findings. Third, as in any systematic review, it may be possible that other relevant instruments have been subjected to rigorous but unreported testing that has not been published may have not been included in this review. To limit this, the authors manually reviewed the grey literature to identify possible observational pain assessment instruments or their testing.

Conclusion

The use of an appropriate observational pain assessment instrument is fundamental to improving pain assessment and management for critically ill intubated adult patients in the ED. This systematic review has highlighted five observational pain assessment instruments for nonverbal critically ill adult patients. While these instruments are commonly used in intensive care to support and optimise pain management, to date, their use in an ED context is absent in the literature. Given an increasing number and range of critically ill patients being managed for prolonged periods of time in the ED setting, a systematic approach to pain assessment is needed for this vulnerable patient group. The review identified that the CPOT had the strongest evidence with regard to validity and reliability for pain assessment in the intensive care critically ill patient. Testing of the CPOT is therefore recommended to determine whether this is the most appropriate instrument for pain assessment in the critically ill intubated patient in the ED. Systematic and consistent use of an appropriate pain assessment instrument in the ED context will optimise pain management and improve patient experiences.

Relevance to clinical practice

Pain is one of the most common presenting complaints in the ED. Effective pain assessment is therefore of prime importance, as it will determine the appropriate type of analgesia required and the urgency of the pain relief needed. Pain management by ED staff is challenging, especially when the patient is critically ill and intubated, and therefore unable to communicate the presence or intensity of acute pain. The use of validated observational pain assessment instruments is needed to guide the assessment, monitoring and administration of analgesia.

Contributions

Conception, design, data extraction and analysis and drafted this article: WV; conception, design and critically revised: MF and DE. All authors have approved this version and agree to be accountable for all aspects of work ensuring integrity and accuracy.

© 2016 John Wiley & Sons Ltd

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

Conflict of interest

All authors declare that they do not have any potential conflicts of interest with respect to the research, authorship and/or publication of this article.

References

Abdi H (2007) Bonferroni and Sidak corrections for multiple comparisons. In *Encyclopedia of Measurement and Statistics* (Salkind N ed.). Sage, Thousand Oaks, CA, pp. 103–107.

Ahlers S, van Gulik L, van der Veen A, van Dongen H, Bruins P, Belitser S, de Boer A, Tibboel D & Knibbe C (2008) Comparison of different pain scoring systems in critically ill patients in a general ICU. *Critical Care* **12**, R15.

Ahlers S, van der Veen A, van Dijk I, Tibboel D & Knibbe C (2010) The use of the Behavioral Pain Scale to assess pain in conscious sedated patients. *Anesthesia and Analgesia* **110**, 127–133.

Aïssaoui Y, Zeggwagh A, Zekraoui A, Abidi K & Abouqal R (2005) Validation of a behavioral pain scale in critically ill, sedated, and mechanically ventilated patients. *Anesthesia and Analgesia* **101**, 1470–1476.

Aitken L, Marshall A, Elliott R & McKinley S (2009) Critical care nurses' decision making: sedation assessment and management in intensive care. *Journal of Critical Nursing* **18**, 36–45.

Anthoine E, Moret L, Regnault A, Sébille V & Hardouin JB (2014) Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes* **12**, 1–10.

Australian Commission on Safety and Quality in Health Care (2013) *Evidence for the Safety and Quality Issues Associated with the Care of Patients with Cognitive Impairment in Acute Care Settings: A Rapid Review*. ACSQHC, Sydney, NSW.

Australian Institute of Health and Welfare (2008). *Australian Hospital Statistics: 2008–2009*. AIHW, Canberra, ACT.

Australian Institute of Health and Welfare (2015) *Emergency Department Care: Australian Hospital Statistics 2014–15*. Health services series no. 65. Cat no. HSE 168. AIHW, Canberra, ACT.

Batiha A-M (2014) Pain management barriers in critical care units: a qualitative study. *International Journal of Advanced Nursing Studies* **3**, 1–5.

Bollen K (1989) *Structural Equations with Latent Variables*. John Wiley & Sons Inc, New York, NY.

Chanques G, Pohlman A, Kress J, Molinari N, de Jong A, Jaber S & Hall J (2014) Psychometric comparison of three behavioural scales for the assessment of pain in critically ill patients unable to self-report. *Critical Care* **18**, R160.

Chen Y-Y, Lai Y-H, Shun S-C, Chi N-H, Tsai P-S & Liao Y-M (2011) The Chinese Behavior Pain Scale for critically ill patients: translation and psychometric testing. *International Journal of Nursing Studies* **48**, 438–448.

Clukey L, Weyant R, Roberts M & Henderson A (2014) Discovery of unexpected pain in intubated and sedated patients. *American Journal of Critical Care* **23**, 216–220.

Critical Appraisal Skills Program UK (2011) *Making Sense of Evidence [Online]*. CASP. Available at: http://www.casp-uk.net/ (accessed 5 May 2016).

Curtis L & Morrell T (2006) Pain management in the emergency department. *Emergency Medicine Practice* **8**, 1–28.

Damström D, Saboonchi F, Sackey P & Bjorling G (2011) A preliminary validation of the Swedish version of the Critical-Care Pain Observation Tool in adults. *Acta Anaesthesiologica Scandinavica* **55**, 379–386.

Desbiens NA, Wu AW, Broste SK, Wenger NS, Connors AF Jr, Lynn J, Yasui Y, Phillips RS & Fulkerson W (1996) Pain and satisfaction with pain control in seriously ill hospitalized adults: findings from the SUPPORT research investigations. For the SUPPORT investigators. Study to understand prognoses and preferences for outcomes and risks of treatment. *Critical Care Medicine*, **24**, 1953–1961.

DeWaters T, Faut-Callahan M, McCann J, Paice J, Fogg L, Hollinger-Smith L, Sikorski K & Stanaitis H (2008) Comparison of self-reported pain and the PAINAD scale in hospitalized cognitively impaired and intact older adults after hip fracture surgery. *Orthopedic Nursing* **27**, 21–28.

Dunwoody C, Krenzischek D, Pasero C, Rathmell J & Polomano R (2008) Assessment, physiological monitoring and consequences of inadequately treated acute pain. *Journal of Perianesthesia Nursing* **23**, S15–S27.

Erstad B, Puntillo K, Gilbert H, Grap M, Li D, Medina J, Pasero C, Varkey B & Sessler C (2009) Pain management principles in the critically ill. *Chest* **135**, 1075–1086.

Gélinas C (2010) Nurses' evaluations of the feasibility and the clinical utility of the Critical-Care Pain Observation Tool. *Pain Management Nursing* **11**, 115–125.

Gélinas C & Johnston C (2007) Pain assessment in the critically ill ventilated adult: validation of the Critical-Care Pain Observation Tool and physiologic indicators. *Clinical Journal of Pain* **23**, 497–505.

Gélinas C, Arbour C, Michaud C, Vaillant F & Desjardins S (2011) Implementation of the critical-care pain observation tool on pain assessment/management nursing practices in an intensive care unit with nonverbal critically ill adults: a before and after study. *International Journal of Nursing Studies* **48**, 1495–1504.

Gélinas C, Fillion L, Puntillo K, Viens C & Fortier M (2006) Validation of the Critical-Care Pain Observation Tool in adult patients. *American Journal of Critical Care* **15**, 420–427.

Gélinas C, Harel F, Fillion L, Puntillo K & Johnston C (2009) Sensitivity and

© 2016 John Wiley & Sons Ltd

specificity of the Critical-Care Pain Observation Tool for the detection of pain in intubated adults after cardiac surgery. *Journal of Pain and Symptom Management* **37**, 58–67.

Georgiou E, Hadjibalassi M, Lambrinou E, Andreou P & Papathanassoglou EDE (2015) The impact of pain assessment on critically ill patients' outcomes: a systematic review. *BioMed Research International* **2015**, 18.

Greenhalgh J, Long A, Brettle A & Grant M (1998) Reviewing and selecting outcome measures for use in routine practice. *Journal of Evaluation in Clinical Practice* **4**, 339–350.

Hawksley H (2000) Pain assessment using a visual analogue scale. *Professional Nurse* **15**, 593–597.

Herr K, Coyne P, Key T, McCaffery M, Manworren R & Merkel S (2006) Pain assessment in the nonverbal patient: position statement with clinical practice recommendations. *Pain Management Nursing* **7**, 44–52.

Herr K, Coyne P, McCaffery M, Manworren R & Merkel S (2011) Pain assessment in the patient unable to self-report: position statement with clinical practice recommendations. *Pain Management Nursing* **12**, 230–250.

Higgins J & Green S (2011) *Cochrane Handbook for Systematic Reviews of Interventions. [Online]*. Available at: http://handbook.cochrane.org/ (accessed 12 January 2016).

Hirschman K, Paik H, Pines J, McCusker C, Naylor M & Hollander J (2011) Cognitive impairment among older adults in the emergency department. *Western Journal of Emergency Medicine* **12**, 56–62.

International Association for the Study of Pain (2006) *Pain Terminology*. IASP, Seattle, WA.

Jacobi J, Fraser GL, Coursin DB, Riker RR, Fontaine D, Wittbrodt ET, Chalfin DB, Masica MF, Bjerke HS, Coplin WM, Crippen DW, Fuchs BD, Kelleher RM, Marik PE, Nasraway SA Jr, Murray MJ, Peruzzi WT, Lumb PD; Task Force of the American College of Critical Care Medicine (ACCM) of the Society of Critical Care Medicine (SCCM), American Society of Health-System Pharmacists (ASHP), American College of Chest Physicians (2002) Clinical practice guidelines for the sustained use of sedatives and analgesics in the critically ill adult. *Critical Care Medicine*, **30**, 119–141.

Jerosch-Herold C (2005) An evidence-based approach to choosing outcome measures: a checklist for the critical appraisal of validity, reliability and responsiveness studies. *The British Journal of Occupational Therapy* **68**, 347–353.

Juarez P, Bach A & Baker M (2010) Comparison of two pain scales for the assessment of pain in the ventilated adult patient. *Dimensions of Critical Care Nursing* **29**, 307–315.

Kabes A, Graves J & Norris J (2009) Further validation of the Nonverbal Pain Scale in intensive care patients. *Critical Care Nurse* **29**, 59–66.

Kanji S, Macphee H, Singh A, Johanson C, Fairbairn J, Lloyd T, Maclean R & Rosenberg E (2016) Validation of the critical care pain observation tool in critically ill patients with delirium: a prospective cohort study. *Critical Care Medicine* **44**, 943–947.

Klien D, Dumpe M, Katz W & Bena J (2010) Pain assessment in the intensive care unit: development and psychometric testing of the Nonverbal Pain Assessment Tool. *Heart and Lung* **39**, 288–297.

Leong I, Chong M & Gibson S (2006) The use of a self-reported pain measure, a nurse-reported pain measure and the PAINAD in nursing home residents with moderate to severe dementia: a validation study. *Age and Ageing* **35**, 252–256.

Leung K, Trevena L & Waters D (2012) Development of an appraisal tool to evaluate strength of an instrument or outcome. *Nurse Researcher* **20**, 13–19.

Li DT & Puntillo K (2006) A pilot study on coexisting symptoms in intensive care patients. *Applied Nursing Research* **19**, 216–219.

Linde S, Badger J, Machan J, Beaudry J, Brucker A, Martin K, Opaluch-Bushy N & Navedo Roy R (2013) Reevaluation of the critical-care pain observation tool in intubated adults after cardiac surgery. *American Journal of Critical Care*, **22**, 491–498.

Ma P, Liu J, Xi X, Du B, Yuan X, Lin H, Wang Y, Su J & Zeng L (2010) Practice of sedation and the perception of discomfort during mechanical ventilation in Chinese intensive care units. *Journal of Critical Care* **25**, 451–457.

Mantha M, Roizen M, Fleisher L, Thisted R & Foss J (2000) Comparison methods of clinical measurement: reporting standards for Bland and Altman analysis. *Anesthesia and Analgesia* **90**, 593–602.

Marmo L & Fowler S (2010) Pain assessment tool in the critically ill post-open heart surgery patient population. *Pain Management Nursing* **11**, 134–140.

Mateo O & Krenzischek D (1992) A pilot study to assess the relationship between behavioral manifestations and self-report of pain in postanesthesia care unit patients. *Journal of Post Anesthesia Nursing* **7**, 15–21.

Merkel S, Voepel-Lewis T, Shayevitz J & Malviya S (1997) The FLACC: a behavioral scale for scoring postoperative pain in young children. *Pediatric Nursing* **23**, 293–297.

Moher D, Liberati A, Tetzlaff J & Altman D (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *British Medical Journal* **339**, 332–336.

Morete M, Mofatto S, Pereira C, Silva A & Odierna M (2014) Translation and cultural adaptation of the Brazilian Portuguese version of the Behavioral Pain Scale. *Brazilian Journal of Intensive Care* **26**, 373–378.

Motov S & Khan A (2009) Problems and barriers of pain management in the emergency department: are we ever going to get better? *Journal of Pain Research* **2**, 5–11.

Mularski R (2004) Pain management in the intensive care unit. *Critical Care Clinics* **20**, 381–401.

National Health and Medical Research Council (2011) *Emergency Care Acute Pain Management Manual*. NHMRC, Canberra, ACT.

Nelson JE, Meier DE, Oei EJ, Nierman DM, Senzel RS, Manfredi PL, Davis SM & Morrison RS (2001) Self-reported symptom experience of critically ill cancer patients receiving intensive care. *Critical Care Medicine* **29**, 277–282.

Nelson JE, Meier DE, Litke A, Natale DA, Siegel RE & Morrison RS (2004) The symptom burden of chronic critical illness. *Critical Care Medicine* **32**, 1527–1534.

O'Connor G, Geary U & Moriarty J (2009) Critical care in the emergency

© 2016 John Wiley & Sons Ltd

department. *European Journal of Emergency Medicine* **16**, 296–300.

Odhner M, Wegman D, Freeland N, Steinmetz A & Ingersoll G (2003) Assessing pain control in nonverbal critically ill adults. *Dimensions of Critical Care Nursing* **22**, 260–267.

Paulson-Conger M, Leske J, Maidl C, Hanson A & Dziadulewicz L (2011) Comparison of two pain assessment tools in nonverbal critically care patients. *Pain Management Nursing* **12**, 218–224.

Payen J-F & Chanques G (2012) Pain assessment in the ICU can improve outcome. *Clinical Pulmonary Medicine* **19**, 21–26.

Payen J-F, Bru O, Bosson J-L, Lagrasta A, Novel E, Deschaux I, Lavagne P & Jacquot C (2001) Assessing pain in critically ill sedated patients by using a behavioral pain scale. *Critical Care Medicine* **29**, 2258–2263.

Pudas-Tähkä S-M, Axelin A, Aantaa R, Lund A & Salanterä S (2014) Translation and cultural adaptation of an objective pain assessment tool for Finnish ICU patients. *Scandinavian Journal of Caring Sciences* **28**, 885–894.

Puntillo K, Morris A, Thompson C, Stanik-Hutt J, White C & Wild L (2001) Patients' perceptions and responses to procedural pain: results from thunder project II. *American Journal of Critical Care* **10**, 238–251.

Puntillo K, Morris A, Thompson C, Stanik-Hutt J, White C & Wild L (2004) Pain behaviors observed during six common procedures: results from Thunder Project II. *Critical Care Medicine* **32**, 421–427.

Puntillo K, Pasero C, Li D, Mularski R, Grap M, Erstad B, Varkey B, Gilbert H, Medina J & Sessler C (2009) Evaluation of pain in ICU patients. *Chest* **135**, 1069–1074.

Richardson D & Mountain D (2009) Myths versus fact in emergency overcrowding and hospital access block. *The Medical Journal of Australia* **190**, 369–374.

Rijkenberg S, Stilma W, Endeman H, Bosman R & Oudemans-Van Straaten H (2015) Pain measurement in mechanically ventilated critically ill patients: behavioral Pain Scale versus Critical-Care Pain Observation Tool. *Journal of Critical Care*, **30**, 167–172.

Rose L, Gray S, Burns K, Atzema C, Kiss A, Worster A, Scales DC, Rubenfeld G & Lee J (2012) Emergency department length of stay for patients requiring mechanical ventilation: a prospective observational study. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* **20**, 30.

Roulin M-J & Ramelet A-S (2012) Pain indicators in brain-injured critical care adults: an integrative review. *Australian Critical Care* **25**, 110–118.

Schug SA, Palmer GM, Scott DA, Halliwell R & Trinca J & APM:SE Working Group of the ANZCA and Faculty of Pain Medicine (2015) *Acute Pain Management: Scientific Evidence*. Australian and New Zealand College of Anaesthetists and Faculty of Pain Medicine, Melbourne.

Schwab D (1980) Construct validity in organization behavior. In *Research in Organizational Behavior* (Staw B & Cummings L eds). JAI Press, Greenwich, CT, pp. 3–43.

Streiner D, Norman G & Cairney J (2015) *Health Measurement Scales: A Practical Guide to their Development and Use*. Oxford University Press, Oxford, UK.

Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, Bouter LM & de Vet HC (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology* **60**, 34–42.

Terwee CB, Mokkink LB, Knol DL, Ostelo R, Bouter LM & de Vet HCW (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research* **21**, 651–657.

Thomas S (2013) Management of pain in the emergency department. *International Scholarly Research Notices: Emergency Medicine* **2013**, 1–19.

Topolovec-Vranic J, Canzian S, Innis J, Pollmann-Mudryj MA, White Mcfarlan A & Baker A (2010) Patient satisfaction and documentation of pain assessments and management after implementing the Adult Nonverbal Pain Scale. *American Journal of Critical Care*, **19**, 345–354.

Topolovec-Vranic J, Gélinas C, Li Y, Pollomann-Mudryi M, Innis J, McFarlan A & Canzian S (2013) Validation and evaluation of two observational pain assessment tools in a trauma and neurosurgical intensive care unit. *Pain Research and Management* **18**, 107–114.

Tripathi T & Dusing S (2015) Long-term neurodevelopmental outcomes of infants born late preterm: a systematic review. *Research and Reports in Neonatology* **2015**, 91–111.

Urden L, Stacy K & Laugh M (2010) *Critical Care Nursing Diagnosis and Management*. Elsevier, St. Louis,MO.

Varndell W, Elliott D & Fry M (2015a) Assessing, monitoring and managing continuous intravenous sedation for critically ill adult patients and implications for emergency nursing practice: a systematic literature review. *Australasian Emergency Nursing Journal* **18**, 1–23.

Varndell W, Elliott D & Fry M (2015b) Emergency nurses' practices in assessing and administering continuous intravenous sedation for critically ill adult patients: a retrospective record review. *International Emergency Nursing* **23**, 81–88.

Varndell W, Fry M & Elliott D (2015c) Emergency nurses' perceptions of sedation management practices for critically ill intubated patients: a qualitative study. *Journal of Clinical Nursing* **24**, 1–10.

Vazquez M, Pardavila M, Lucia M, Aguado Y, Margall M & Asiain M (2011) Pain assessment in turning procedures for patients with invasive mechanical ventilation. *Nursing Critical Care* **16**, 178–185.

Voepel-Lewis T, Zanotti J, Dammeyer J & Merkel S (2010) Reliability and validity of the face, legs, activity, cry, consolability behavioral tool in assessing acute pain in critically ill patients. *American Journal of Critical Care* **19**, 55–61.

Warden V, Hurley A & Volicer L (2003) Development and psychometric evaluation of the Pain Assessment in Advanced Dementia (PAINAD) scale. *Journal of the American Medical Directors Association* **4**, 9–15.

Weir S & O'Neill A (2008) Experiences of intensive care nurses assessing sedation/agitation in critically ill patients. *Nursing in Critical Care* **13**, 185–194.

Young J, Siffeet J, Nikoletti S & Shaw T (2006) Use of a Behavioural Pain Scale to assess pain in ventilated, unconscious and/or sedated patients. *Intensive and Critical Care Nursing* **22**, 32–39.

© 2016 John Wiley & Sons Ltd