

RESEARCH

Open Access



Pain assessment tools for use in infants: a meta-review

Diana Arabiat^{1,2*}, Evalotte Mörelius^{2,3}, Kreshnik Hoti⁴ and Jeffery Hughes⁵

Abstract

Background Identifying pain in infants is challenging due to their inability to self-report pain, therefore the availability of valid and reliable means of assessing pain is critical.

Objective This meta-review sought to identify evidence that could guide the selection of appropriate tools in this vulnerable population.

Methods We searched Scopus, Medline, Embase, CINAHL, MIDRIS, EMCare and Google Scholar for eligible systematic reviews. Eligible reviews documented psychometric properties of available observational tools used to assess pain in infants.

Results A total of 516 reviews were identified of which 11 met our inclusion criteria. We identified 36 pain assessment tools (evaluated in 11 reviews) of which seven were reported in at least three reviews. The level of evidence reported on the psychometric properties of pain assessment tools varied widely ranging from low to good reliability and validity, whilst there are limited data on usability and clinical utility.

Conclusions Currently, no observer administered pain assessment tool can be recommended as the gold standard due to limited availability and quality of the evidence that supports their validity, reliability and clinical utility. This meta-review attempts to collate the available evidence to assist clinicians to decide on what is the most appropriate tool to use in their clinical practice setting. It is important that researchers adopt a standard approach to evaluating the psychometric properties of pain assessment tools and evaluations of the clinical utility in order that the highest level of evidence can be used to guide tool selection.

Keywords Infants, Validity, Reliability, Clinical utility, Observational pain assessment tools

Background

Pain occurs across the continuum of life. However, for those who cannot self-report their pain, such as pre-verbal children, those with intellectual disability and those with neurological disorders who have lost the ability to communicate (e.g., people living with advanced dementia), there is evidence that their pain often goes under-recognized and under-treated [1–3]. Research clearly indicates that accurate pain assessment in infants is a challenge for health care professionals [2, 4]. The problem of differentiating or discriminating pain from other expressions of unpleasant feelings or experiences, such as

*Correspondence:

Diana Arabiat

d.arabiat@ecu.edu.au; d.arabiat@ju.edu.jo

¹ Maternal and Child Nursing Department, Faculty of Nursing, The University of Jordan, Amman, Jordan

² School of Nursing and Midwifery, Edith Cowan University, Joondalup, Western Australia, Australia

³ Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

⁴ Faculty of Medicine, University of Prishtina, Prishtina, Kosovo

⁵ Curtin Medical School, Faculty of Health Sciences, Curtin University, Bentley, WA, Australia



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

hunger, fear or distress adds to the complexity of assessing pain in this group [5].

In the absence of infants' ability to self-report, observation of behaviours such as facial expression, crying, hand and leg movements is a valid approach for pain assessment [3, 4, 6]. However, pain behaviours are not specific reflections of pain intensity or distress, and the context of infant's behaviour must be considered when using this approach [6].

Herr et al. [6], report that typically children as young as 3 years may be able to report pain using basic tools, however it is not until 8 years of age that children are generally considered reliable in using a numerical rating scale. In clinical practice, Eccleston et al. [7], in support of Beltramini et al. [8], suggest that self-reported measures of pain (e.g., numerical scales) can be used for children aged 6 years and older, while behavioral scales can be used for children aged below 6 years or those unable to verbally report. According to the American Society of Pain Management Nursing (APSMN) position statement on pain assessment in the patient unable to self-report, health care professionals should search for potential causes of pain and use both primary behavioral categories and behavioral pain tools for initial and ongoing assessment of pain [6].

An essential element in assessment and management of pain is the identification of the type and source of pain, both of which can be challenging in the infant population [8]. The following types of pain have been identified in infants: acute pain that occurs in response to recent illness, injury or following procedures or operations, and has a limited duration; and chronic pain which commonly persists beyond the time of curing of an injury or illness, and often lasts beyond 6 months [9]. Validated assessment tools in infants have been developed mostly within research into acute pain. These included comprehensive identification of primary behavioral categories, such as facial expression, body activity/motor movement, and crying/verbalization [2]. For assessment of chronic pain, infants may demonstrate different behaviours, such as withdrawal, lack of expression, lack of interest in surroundings, and a decrease in their ability or their willingness to play [10].

Eccleston et al. [7] proposed four transformative goals to improve the lives of children and adolescents with pain and their families. These were to 1) "make pain matter", 2) "make pain understood", 3) "make pain visible", and 4) "make pain better". They pointed out that whilst on face value these goals might seem obvious and simple, the ongoing occurrence of acute pain, pain post-surgery and procedures, and chronic pain in this population suggest otherwise. Considering the above goals highlighted by Eccleston et al. [7] of particular importance remains the

issue of adequately identifying and therefore assessing pain using evidence-based tools. In this regard, previous reviews have identified that while there are a number of tools that can be used in infant population, the strength of psychometric evidence to support their use varies [8, 11]. Therefore, there is an ongoing need to review evidence and therefore progress made in this area. Meta-reviews, also known as systematic reviews of systematic reviews, provide a rigorous approach for synthesizing evidence from available systematic reviews and summarizing knowledge into one accessible document [12]. Given the wide variety of pain assessment tools and uncertainty as to which assessment tools are more suitable when used in clinical or multiple settings and within different populations, we felt that a meta-review was particularly useful for our purpose. Undertaking this meta-review would provide an updated systematic identification and appraisal of the evidence concerning the validity and reliability of pain assessment tools available for use in infants aged 1-12 months old. This would then allow an evaluation of the appropriateness of these pain tools for use in different clinical settings. This age range was selected in order to exclude tools which are used exclusively for pre-term and full-term neonates.

Methods

Design

We performed a systematic literature review of published reviews in accordance with a protocol that was published in advance on the PROSPERO database with reference number CRD42021236227. We used the Joanna Briggs Institute (JBI) methodology for systematic review of systematic reviews [12] and the Preferred Reporting Items for Systematic Review [13] and Meta-Analysis (PRISMA-s) guideline for conducting and reporting our findings [14].

Review question and eligibility

This meta-review aimed to address the following questions: 1) What pain assessment tools are available for assessing pain in infants; 2) What are the documented psychometric properties of the available tools in this population; and 3) How have previous reviews evaluated and recommended these pain tools for use in infants?

Studies needed to meet the following inclusion criteria to be eligible for inclusion in this review (the formulation of the research question with reference to PICO (Patient/Population, Intervention, Comparison and Outcomes) is provided in the Supplementary Table 1):

- 1) Participants: This meta-review considered reviews that report on the pain assessment tools used for infants, of either gender. Infant in this review is

defined as any child aged between 1 and 12 months of age. Children less than 1 month were considered neonates [15], and tools used exclusively to assess this population were excluded.

- 2) Exposure of interest: The exposure of interest is pain assessment tools. Reviews that include management of pain were considered if they have also covered assessment of pain. The International Association for the Study of Pain (IASP) suggested that pain assessments entail a comprehensive evaluation of the patient's pain, symptoms, functional status, and clinical history [16]. According to Stevens et al. [17], acute pain is often limited to a short period of time and in infants is frequently associated with procedures (e.g. venipuncture, immunization, wound dressings), trauma (e.g. burns), or post-surgery, while chronic pain persists beyond 3 – 6 months and may either be recurrent as is the case with headache, backache, stomach aches or prolonged as that associated with use of ventilators for an extended period of time. For this review, all forms of pain were considered.
- 3) Outcomes: This meta-review considered reviews that included pain assessment tools irrespective of the outcomes of the assessment (e.g., infant being in pain or not).
- 4) Types of studies: This meta-review considered all type of reviews including integrative, scoping, and meta-analysis studies that have evaluated the validity and reliability of pain assessment tools used in infants.

Reviews of studies of the assessment of pain and pain assessment tools were regarded as valid for the purpose of this meta-review if the conditions listed below were met: (1) the review carried out systematically (e.g., publication that makes explicit the authors' intention to review or summarize the literature; with review, overview, or meta-analysis in the title or in the section heading); (2) entails a clear set of objectives (explicit and clear research question); (3) reproducible methodology (the paper clearly explains how the evidence was retrieved, including sources and search strategy, and the inclusion and exclusion criteria); (4) satisfy a clear assessment of validity of the findings (e.g., assessment of risk of bias); (5) satisfy a systematic presentation and synthesis of findings beyond those provided by single studies; and (6) reviews of pain assessment tools that include other populations such as premature neonates, toddlers, adolescents or adults were also included provided they have included pain assessment tools used for infants.

The current meta-review excluded reviews with unclear design method, reviews limited to neonates and premature neonates, reviews of systematic reviews,

and studies published in languages other than English. To account for the complexity with the use of facial expression in premature neonates [18] and the distinct difference in pain expression in premature neonates and full-term neonates less than 1 month [19, 20], we decided to exclude pain assessment tools exclusively used for this population.

Datasourcesandsearchstrategy

Articleselection

We searched Scopus, Medline, Embase, CINAHL, MIDRIS, EMCare and Google Scholar (Supplementary Table 2 details databases searched, dates of searches and outcomes, Supplementary Table 3 details the keywords used in the search and Supplementary Table 4 search strategies and outputs for each database searched). EndNote was used to keep the results of the searches of each database and then they were entered into Rayyan, a web-based tool for systematic reviews [21]. After removal of duplicates, all abstracts were initially screened against eligibility criteria by two reviewers in a blinded standardized manner (DA and MA), and then all selected articles were read in full text to receive the final inclusion. Any disagreement between the two reviewers regarding the eligibility of any studies was resolved through discussion and inclusion of a third reviewer (EM).

Dataextractionandanalysis

For each review included, we extracted descriptive information on study characteristics and study results using a structured data table. Information which was obtained included study authors and year of publication, country, type of review, name of pain assessment tool recommended for infants and type of psychometric data obtained. A list of major definitions of psychometric properties used for extracting data is presented in Table 1. The latter domain also contained the JBI Risk of Bias Assessment Tool for evaluation of study quality. Two reviewers independently extracted data and rated the risk of bias for each of the included studies (DA, MA) and resolved discrepancies through team discussion. Then, the results extracted from the reviews were synthesized using a narrative synthesis approach. Data synthesis was conducted separately for each of the eligible pain assessment tools listed in the review, including general recommendations for using the tool for infants and its psychometric properties. The data extracted for this section was in accordance with the definitions given by the authors in their review and their evaluation of type of measurement property obtained.

Table 1 Definitions of psychometric properties and subcategories. (All reliability and validity definitions were extracted and adopted from Walsh & Betz [22])

Property	Definition
Reliabilityassessment	It is the degree to which the measuring of an attribute by a tool is systematic and therefore repeatable.
Internal consistency	It is the degree to which each item of a tool is measuring the same thing as each other item (Cronbach's alpha, $\alpha \geq 0.70$ for group comparisons)
Inter-rater reliability	It is the correlation between ratings of the same rater at two different times, or the product moment correlation between ratings of two different ratters using the same tool (Intraclass correlation coefficient ICC recommended or Pearson's $r \geq 0.70$ for group comparisons, Fleiss' Kapp, or percent agreement)
Validityassessment	It is the extent to which the tool actually measures the characteristic or dimension that it is intended to measure.
Face validity	It is the degree to which the content of a test appears relevant to the concept that the tool is intended to measure (judged by a group of experts)
Content validity	It is the degree to which an assessment tool holds an appropriate list of items to represent the concept of interest (basic and minimum index of content validity).
Convergent validity	It is the evidence that different assessment tools developed to measure the same concept all measure the same trait (Pearson's recommended ≥ 0.40)
Discriminant validity	It is the degree of dissimilarity or distinctness of a tool scores that theoretically represent different trait
Cross cultural validity	It is the degree to which the performance of the assessment tool when being translated or culturally adapted for another population or settings is an adequate reflection of the psychometric performance of the original tool.
Criterion related validity	It is the ability to test if an assessment tool is able to predict a variable that is designated as a criterion or not. It is often measuring the correlation of the instrument with a "gold standard" criterion administered at the same time. It includes two forms (1) predictive validity and (2) Concurrent validity
Construct validity	It is the extent to which a tool measures the construct it is designed to measure and in the settings that it is designed for (factor analysis including exploratory and confirmatory factor analysis)
Clinicalutility	It is the degree to which actual use of the assessment tool in clinical settings is associated with changing health outcomes
Clinicalfeasibility	It is the practical extent to which an assessment tool can be plausible in a given population or clinical setting

Risk of bias and methodological quality

The 10 reviews that met our inclusion criteria were assessed for methodological quality by two reviewers independently (DA, MA) using the JBI Critical Appraisal Checklist for Systematic Reviews and Research Syntheses [23]. Disagreements between the two reviewers were resolved through discussion and if needed involvement by a third reviewer (EM).

Quality assessment according to the JBI checklist tool involved 11 assessment criteria: clarity and explicitly of the review questions, inclusion criteria, search strategy, adequacy of sources and resources used, criteria for study appraisal, number of reviewers (2 or more), methods to minimize errors in data extraction, methods used for combined studies, assessment of publication bias, recommendation for policy and practice, and direction for new research (Table 2). Every criterion in the checklist was given a rating of 'yes', 'no', 'unclear' or 'not applicable,' and one point was given to every criterion rated 'yes.' Using the JBI checklist tool, methodological quality can be judged in terms of "low" if they failed to reach a score of > 50% on critical appraisal, the predetermined cut off score agreed upon by the research team.

Results

Selectionofarticles

The literature search produced 409 potentially relevant reviews for screening of which 112 were duplicates, leaving 297 unique reviews. After title and abstract screening 45 were included in the full text review. After full text review, 35 studies were excluded (See Supplementary Table 5), providing a total of 10 studies (See Supplementary Table 6) for data extraction (Figure 1). An updated database search was conducted on the 1st of April 2023 to identify the latest research on the topic. The search yielded 229 studies (Addendum 1). These studies were screened for eligibility as presented in Figure 1. Only one of the studies was identified as eligible. Figure 1 below presents a schema for the search process and the outcome of the article selection process.

Characteristicsofincludedreviews

This meta-review included one systematic integrative review [30] and 10 systematic reviews [24–29, 31–34]. The number of studies included in each review varied from 6 to 250 studies although the number of included studies in some reviews was unclear or not provided (Table 3). The reviews were published between the years

Table 2 Critical appraisal studies of reviews

Author(year)	Q.1	Q.2	Q.3	Q.4	Q.5	Q.6	Q.7	Q.8	Q.9	Q.10	Q.11	Totalscore
Chan,etal.(2022) [24]	Y	Y	Y	Y	Y	Y	Y	UN	N	Y	Y	9/11
Bai&Jiang(2015) [25]	Y	Y	Y	Y	Y	Y	Y	UN	N	Y	UN	8/11
Crellinetal.(2007) [26]	Y	Y	Y	Y	NA	N	N	N	N	Y	Y	6/11
Crellinetal.(2015) [27]	Y	Y	Y	Y	Y	Y	Y	NA	N	Y	Y	9/11
Crellinetal.(2018) [28]	Y	Y	Y	Y	Y	Y	Y	NA	Y	Y	Y	10/11
Crostaetal.(2014) [29]	Y	Y	Y	Y	UN	N	N	UN	N	Y	UN	5/11
Duhn&Medves(2004) [30]	Y	Y	Y	Y	N	N	N	N	N	Y	Y	6/11
Giordanoetal.(2019) [31]	Y	Y	Y	Y	Y	Y	Y	NA	Y	Y	Y	10/11
Stapelkampetal.(2011) [32]	Y	Y	Y	Y	Y	Y	Y	NA	N	Y	Y	9/11
Maaskantetal.(2016) [33]	Y	Y	Y	Y	Y	Y	Y	NA	Y	Y	Y	10/11
Kingsnorthetal.(2015) [34]	Y	Y	Y	Y	Y	Y	Y	UN	N	Y	Y	9/11

Q.1 Questions are clearly and explicitly stated; Q.2 Inclusion criteria are appropriate for the review question; Q.3 The search strategy is appropriate; Q.4 The sources and resources used to search for studies are adequate; Q.5 The criteria for appraising studies are appropriate; Q.6 Critical appraisal was conducted by two or more reviewers independently; Q.7 There were methods to minimize errors in data extraction; Q.8 The methods used to combine studies are appropriate; Q.9 The likelihood of publication bias was assessed; Q.10 Recommendations for policy and/or practice are supported by the reported data; Q.11 The specific directives for new research are appropriate

Y Yes, N No, NA Not applicable, UN Unclear

2004 and 2022 and they reported on between 1 to 29 pain assessment tools which can be used for infants aged 1 to 12 months (Table 3).

Synthesis and findings

We identified 11 reviews that contained tools measuring pain in infants (1-12 months) and provided psychometrics analysis of 36 tools. Two tools were assessed in five reviews: the FLACC (Face, Legs, Activity, Cry and Consolability) [25–27, 31, 32] and the COMFORT/COMFORT-behaviour [24, 25, 31–34], while two were included in four reviews, namely the CHEOPS (The Children Hospital of Eastern Ontario Pain Scale) [25, 30, 32, 33] and the PPP (Paediatric Pain Profile) [24, 29, 32, 34]. The MBPS/MBPS-R (Modified Behavioral Pain Scale/Modified Behavioral Pain Scale Revised) [28, 30, 31], the CSS/POPS (Clinical Scoring System also referred to as the Post-Operative Pain Score) [30–32], and the TPPPS (Toddler Pre-schooler Poster-Operative Pain Scale) [26, 31, 32] were included in three reviews. While the CHIPPS (Children’s & Infant’s Postoperative Pain Scale) [29, 31], the PEPP/mPEPP (Preverbal, Early verbal Pediatric Pain Scale/modified Preverbal, Early verbal Pediatric Pain Scale) [26, 31]; and RIPS (Riley Infant Pain Scale) [30, 31] were included in two reviews. The remainder of the tools were assessed only by one review.

Settingswherethetoolswerestudied

The child population varied across reviews and some reviews included mixed population of newborns and infants [30, 32, 33] infants and toddlers [30, 32], infants and older children or adolescents [24, 27, 28, 31, 32, 34]

. Two reviews were limited to children with cognitive impairments [29], or cerebral palsy [34]. Twenty-two tools focused on pre-, peri- and/or post-operative pain, 10 tools focused on procedural pain and 10 on acute pain, while the rest were not adequately described. The terms used to describe the use in hospital settings included bedside, Paediatric Intensive Care Unit (PICU), intensive care, ventilated child, sedated child, post analgesic, and pre-operative, peri-operative and post-operative. The terms used to describe the use for acute procedural settings also varied and included procedural/ immunization/ outpatient clinic, triage in Accident & Emergency or procedural outpatient.

Assessmentofevidenceofmeasurementpropertiesofthetools

We described the detailed study characteristics in Table 3 and the measurement properties in Table 4. The quality of retrieved reviews was assessed at the level of data synthesis, concerning the quality of methodology. Five reviews used the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist [24, 27, 28, 31, 33] one review used the Psychometric Property Coding System – CODE [25] and one review used the Scottish Intercollegiate Guidelines Network (SIGN evidence-grading system) [32]. The authors of one review critiqued tools on the merit of their validity and reliability using the rating schema for assessing clinical utility developed by the Society of Paediatric Psychology Assessment Task Force [34], while the rest of retrieved reviews did not describe their rating system [26, 29, 30]. It is worth noting here that there are differences between

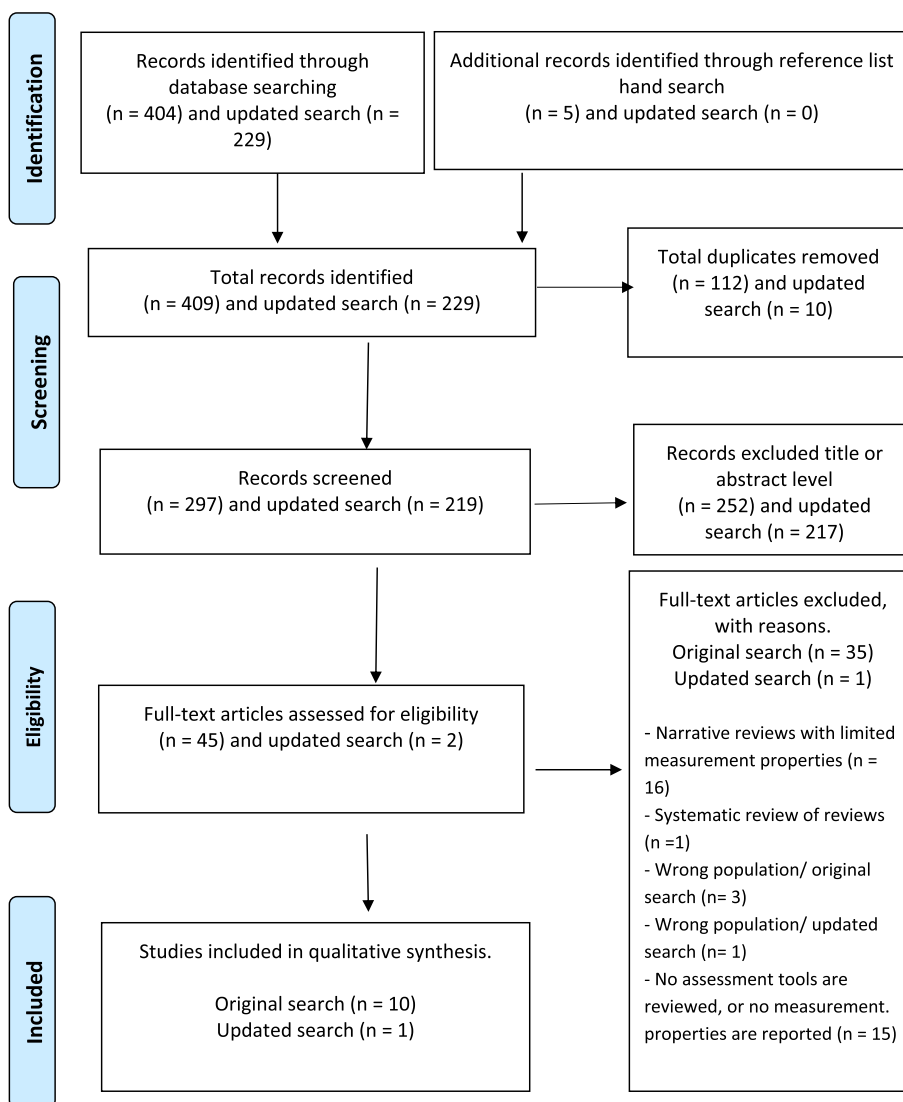


Fig. 1 PRISMA flow diagram of records identified in the search of eligible reviews

the assessments of evidence of measurement properties used between studies. For example, COSMIN [35] is used to evaluate measurement properties of a scale (scale level), and SIGN [36] is used to evaluate the overall methodological quality of primary measurement studies. On the other hand, the original psychometric properties coding system for evaluating methodological quality of the assessment scales developed by Zwakhalen et al. [37] was adopted and refined for use by Bai and Jiang [25]. The differences between the reviews in terms of their overall aim, inclusion criteria, and their rating system made it difficult to aggregate the results and make a conclusion about the psychometric properties of the tools included in this review.

Psychometric data of the pain assessment tools in infant

The 36 pain assessment tools were assessed for properties such as internal consistency (18 tools), content/face validity (30 tools), criterion validity (23 tools), construct validity (23 tools), and cross-cultural validity (3 tools). Overall, and as shown in Table 4, the methodological quality reported in the reviews varied significantly and ranged from ‘acceptable’ for the POCIS [25] to very good’ for the FLACC, COMFORT-Behavior (COMFORT-B), and POCIS [25], and from high risk of bias for the Nursing Assessment of Pain Intensity Revised (NAPI), POCIS, Post-Operative Pain Score (POPS), Riley Infant Pain Scale (RIPS), and University of Wisconsin Children’s Hospital Pain Scale (UWCH) to low risk of bias for the

Table 3 Details of the included reviews and the eligible pain assessment tools they included

Author (year) Country	Review type	Aim of the review	Total number of studies included	Pain assessment tools reviewed related to infants	Ages included	Setting	Risk of bias
Chan, et al. (2022) [24] China/ United Kingdom	Systematic review	To identify pain assessment tools used in palliative care of children	34	DOLIS	0-11 mo	Procedural in cancer	The COSMIN checklist
				COMFORT	0 – 3 yr	Post-operative	
				DEGR	0 – 3 yr	Acute pain	
Bai & Jiang (2015) [25] United States of America/China	Systematic review	To assess the psychometric properties of available assessment tools for measuring pain in Chinese children.	6	MIPS	1 – 7 mo	Post-operative	
				PPP	1 – 18 yr	Neurological and cognitive impairment	
				FLACC	0-7 yr	Post-operative, Procedural in burn	Psychometric property coding system- CODE
Crellin, et al. (2007) [26] Australia	Systematic review	To identify and critically analyze the validity and reliability of preverbal and early verbal behavioral pain assessment tools.	N/A	COMFORT-B	0-7 yr	Post-operative,	
				POCIS	0-5 yr	Procedural pain	
				CHEOPS	6-12 mo	Post-operative	Not described
Crellin, et al. (2015) [27] Australia	Systematic review	To evaluate the psychometric properties and utility of the FLACC scale for use in research and to evaluate its appropriateness for clinical and research settings	77	FLACC	< 3 yr	Post-operative	
				TPPPS	1-5 yr	Post-operative, Acute pain	
				PEPPS	12-24 mo	Post-operative	
				FLACC	6 mo – 4 yr	Post-operative, Procedural/immunization pain, PICU	The COSMIN checklist

Table 3 (continued)

Author (year) Country	Review type	Aim of the review	Total number of studies included	Pain assessment tools reviewed related to infants	Ages included	Setting	Risk of bias
Crellin, et al. (2018) [28] Australia	Systematic review	To assess the research evidence addressing the validity and reliability of the Modified Behavioral Pain scale and its appropriateness for use in children	28	MBPS	0–18 yr	Post-operative, Procedural/immunization pain	The COSMIN checklist
Crosta, et al. (2014) [29] United States of America	Systematic review	The aims of this review were To evaluate the best available evidence for use of pain assessment tools in hospitalized children with cognitive impairment and to examine their clinical utility in acute care settings.	45 studies	PPP	1–18 yr,	Multi-setting; home, postop. Inpatient. Nonverbal children with cognitive impairment	Not described
Duhn & Medves. (2004) [30] Canada	Systematic integrative review	To examine the validity and reliability of published pain assessment tools in infants' tools and to evaluate their reported clinical utility and feasibility	N/A Total of 35 infant pain assessment tools	PRS RIPS MAX CHEOPS CSS or POPS BPS MBPS CHIPPS	1–12 mo Infants & children 0–2 yr 1–5 yr 1–7 mo 3–36 mo 4–6 mo Newborns, infants & young children	Not described Post-operative Not described Post-operative Post-operative Ventilated child Not described Not described	Not described

Table 3 (continued)

Author (year) Country	Review type	Aim of the review	Total number of studies included	Pain assessment tools reviewed related to infants	Ages included	Setting	Risk of bias
Giordano, et al. (2019) [31] Austria	Systematic review	To provide clinicians with a complete overview on the validity and reliability of the existing pain and sedation scales for different target populations (preterm infants, term infants, and toddlers) and in different clinical contexts.	89 studies	AHTPS	0-16 yr	Acute pain	The COSMIN checklist
				BOPS	1-7 yr	Post-operative	
				CAAS	0-16 yr	Post-operative, Intubated, Ventilated patient	
				CHEOPS	1-5 yr	Post-operative	
				CHIPPS	0-5 yr	Post-operative	
				COMFORT	0-17 yr	Post-operative	
				COMFORT-B	0-3 yr	Pre-post-surgery	
				EVENDOL	0-7 yr	Acute and prolonged pain	
				FLACC	1 d-7 yr	Acute pain, Post-operative	
				FPS-R	0-3 mo	Post-operative	
				Hartwig Scale	0-10 mo	Procedure	
				MAPS	0-31 mo	Post-surgery	
				MBPS	4-6 mo	Procedure	
				MIPS	1-7 mo	Pre-post-operative	
				MBPS-R	2-12 mo	NA	
				NFCS-R	2-12 mo	Procedure	
				NIPS	2-6 mo	Procedure	
				NAPI	0-36 mo	Post-operative	

Table 3 (continued)

Author (year) Country	Review type	Aim of the review	Total number of studies included	Pain assessment tools reviewed related to infants	Ages included	Setting	Risk of bias
Kingsnorth, et al. (2015) [34] Canada	Systematic review	To identify and critique pain assessment tools used for measuring chronic pain in children and evaluate its appropriateness for use in children with cerebral palsy.	250	ObsVAS	0-4 yr	Post-surgery	Rating schema for assessing validity, reliability and clinical utility
				POCIS	0-4 yr	Burned child, Post-operative, Acute pain	
				P-MIPS	1-7 mo	Pre- Post-surgery	
				POPS	0-36 mo	Post-operative	
				PEPPS	12-24 mo	Post-operative	
				mPEPPS	12-84mo	Procedure	
				RIPS	0-36 mo	Post-operative	
				RCEMCPS	0-16 yr	Acute pain	
				TPPPS	12-64 mo	Post-operative	
				TVP	0-13 mo	Acute pain	
				UWCH	0-16 yr	Acute pain	
				PPP	1-18 yr	Children with cerebral palsy	
				Stapelkamp, et al. (2011) [32] United Kingdom	Systematic review	To review and update the best available evidence on pain assessment in children.	
CAAS	<3 yr ^a	PICU post-operative					
CMPPMS	<3 yr ^a	Post-operative					
CHEOPS	<3 yr ^a	Post- and peri-operative					
COMFORT	<3 yr ^a	Post- and peri-operative					
DCHPT	<3 yr ^a	Post-operative					
FLACC	<3 yr ^a	Post-operative and peri-procedural					
NAPI	<3 yr ^a						
POPS	<3 yr ^a	Post-operative					
PPP	1-18 yr	All settings					

Table 3 (continued)

Author (year) Country	Review type	Aim of the review	Total number of studies included	Pain assessment tools reviewed related to infants	Ages included	Setting	Risk of bias
Maaskant, et al. 2016 [33] Netherlands	Systematic review	To evaluate the methodological quality of studies used the COMFORT scale in children up to 18 years.	30	PRS TPPPS UWCH ObsVAS COMFORT	<3 yr ^a <3 yr ^a <3 yr ^a <3 yr ^a 0–18 yr	Post-operative Post-operative Peri-procedural pain Post-operative pain PICU, Non ventilated neonates, NICU	The COSMIN checklist

AHTPS: Alder Hey Triage Pain Scale, *BOPS*: Behavioural Observational Pain Scale, *BPS*: Behavioural Pain Score, *CAAS*: Cardiac Analgesic Assessment Scale, *CHEOPS*: The Children Hospital of Eastern Ontario Pain Scale, *CHIPPS*: Children's and Infants Postoperative Pain Scale, *CHIPPS* or *KUSS*: Children's & Infant's Postoperative Pain Scale, *CMPPMS*: Chedoke-McMaster Paediatric Pain Management Sheet, *DCHPT*: Derbyshire Children's Hospital Pain Tool, *DEGR*: Douleur Enfant Gustave Roussy Scale, *FLACC*: Face, Legs, Activity, Cry and Consolability, *EVENDOL*: Evaluation Enfant Douleur, *FPS-R*: Faces Pain Scale Revised, *MAPS*: Multidimensional Assessment Of Pain Scale, *MAX*: Maximally Discriminative Facial Movement Coding System, *MBPS*: Modified Behavioural Pain Scale, *MIPS*: Modified Infant Pain Scale, *mo*: months, *MAPI*: Nursing Assessment of Pain Intensity Revised, *MFCSr*: Neonatal Facial Coding System-Revised, *NIPS*: Neonatal Infant Pain Scale, *ObsVAS*: Observer Visual Analog Scale, *PEPPS*: Preverbal, Early verbal Paediatric Pain Scale, *P-MIPS*: Partial Modified Infant Pain Scale, *POCIS*: Pain Observation Scale for Young Children, *POPS*: Postoperative Pain Score, *PPP*: Paediatric Pain Profile, *PRS*: Pain Rating Scale, *RCEMCPs*: Royal College of Emergency Medicine Composite Pain Scale, *RIPS*: Riley Infant Pain Scale, *TPPPS*: Toddler Pre-schooler Postoperative Pain Score, *TVP*: Touch Visual Scale, *UWCH*: University of Wisconsin Children's Hospital Pain Scale, yr: years

^a Age < 3 yr excludes neonates

Table 4 Measurement properties of pain assessment tools for infants extracted included reviews

Author (year)	Assessment tool	Reliability assessment			Validity assessment	
		Internal consistency	Inter-rater reliability	Intra-rater reliability	Face/Content validity	Convergent validity
Chan, et al. (2022) [24] ^a	COMFORT	High	High	-	Moderate	-
	DOLLS	High	Moderate	-	-	-
	FLACC	Moderate	High	-	-	-
	MIPS	-	Very low	-	-	-
	PPP	High	Moderate	-	Very low	-
Bai & Jiang. (2015) [25] ^b	FLACC	$\alpha=0.75 - 0.799$	ICC=0.79 - 0.84 $r=0.65 - 0.95$	-	Yes	Acceptable to high
	COMFORT- behavior	$\alpha=0.869 - 0.83$	ICC=0.82 - 0.9	-	Yes	-
	POCIS	$\alpha=0.846 - 0.856$	ICC=0.78 - 0.80, $r=0.66$	-	Yes	-
Crellin, et al. (2007) [26] ^a	CHEOPS	Moderate to good	Very good	-	-	-
	FLACC	-	Good	-	-	Demonstrated
	TPPS	Poor to Good	Good	-	-	-
	PEPPS	Poor to good	Very good	Very good	-	-
Crellin, et al. (2015) [27] ^a	FLACC	Moderate	Low, Moderate, High	Moderate to High	Low	Very low to moderate
Crellin, et al. (2018) [28] ^a	MBPS	Low to Moderate	Low, Moderate, High	Moderate to High	-	Nil to Low
Crosta, et al. (2014) [29] ^(d)	PPP	$\alpha=0.75 - 0.89$	ICC=0.70 - 0.89	-	Yes	-
Duhn & Medves. (2004) [30] ^a	PRS	-	$r=0.65$ to 0.84, $P=0.0001$	-	-	-
	RIPS	$\alpha=0.87$ to 0.93	ICC=0.39-0.87	-	-	-
	MAX	-	Agreement83%	-	-	-
	CHEOPS	-	Agreement90 to 99.5%	-	Yes	$r=0.85, p<0.05$
	CSS or POPS	-	-	-	-	-
	BPS	-	-	-	-	-
	MBPS	-	ICC=0.95, $p<0.001$	-	-	-
	CHIPPS	$\alpha=0.96$	$r=0.93$	-	-	-

Table 4 (continued)

Author (year)	Assessment tool	Reliability assessment			Validity assessment	
		Internal consistency	Inter-rater reliability	Intra-rater reliability	Face/Content validity	Convergent validity
Giordano et al. (2019) [31]^a	AHTPS	-	Demonstrated	-	Yes	-
	CAAS	-	Demonstrated	-	Yes	-
	BOPS	-	Demonstrated	-	Yes	-
	CHEOPS	Demonstrated	Demonstrated	-	Yes	Established (with VAS)
	CHIPPS	Demonstrated	Demonstrated	-	Yes	-
	COMFORT	Demonstrated	Demonstrated	-	Yes	-
	COMFORT-B	-	Demonstrated	-	Yes	-
	EVENDOL	-	Demonstrated	Demonstrated	Yes	-
	FLACC	Demonstrated	Demonstrated	Demonstrated	Yes	-
	FPS-R	-	Demonstrated	-	Yes	-
	Hartwig	Demonstrated	Demonstrated	-	Yes	Established
	MAPS	Demonstrated	-	-	Yes	-
	MBPS	Demonstrated	Demonstrated	Demonstrated	Yes	-
	MBPSr	Demonstrated	-	Demonstrated	Yes	-
	MIPS	-	-	-	Yes	-
	NFCSr	Demonstrated	-	-	Yes	-
	NIPS	Demonstrated	Demonstrated	-	Yes	-
	NAPI	Demonstrated	Demonstrated	-	Yes	-
	ObsVAS	-	Demonstrated	-	Yes	-
	POCIS	Demonstrated	Demonstrated	-	Yes	-
	PMIPS	-	-	Demonstrated	Yes	-
	POPS	-	-	-	Yes	-
	PEPPS	Demonstrated	Demonstrated	Demonstrated	Yes	-
mPEPPS	Demonstrated	Demonstrated	-	Yes	-	
RIPS	-	Demonstrated	Demonstrated	Yes	-	
RCEMCPS	-	-	Demonstrated	Yes	-	
TPPPS	Demonstrated	Demonstrated	-	Yes	-	
TVP	-	-	-	Yes	-	
UWCH	-	Demonstrated	Demonstrated	Yes	-	
Kingsnorth et al. (2015) [34]^f	PPP	Well-established	-	-	-	-
Stapelkamp et al. (2011) [32]^c	AHTPS	-	Acceptable	-	-	-
	CASS	-	Acceptable	-	-	-
	CMPPMS	-	Acceptable	-	-	-
	CHEOPS	-	Acceptable	-	-	-
	COMFORT	-	Acceptable	-	-	-
	DCHPT	-	Acceptable	-	-	-
FLACC	-	Acceptable	-	-	-	

Table 4 (continued)

Author (year)	Assessment tool	Reliability assessment			Validity assessment	
		Internal consistency	Inter-rater reliability	Intra-rater reliability	Face/Content validity	Convergent validity
	NAPI	-	Acceptable	-	-	-
	POPS	-	Acceptable	-	-	-
	PPP	-	Acceptable	-	-	-
	PRS	-	Acceptable	-	-	-
	TPPPS	-	Acceptable	-	-	-
	UWCH	-	Acceptable	-	-	-
	ObsVAS	-	Acceptable	-	-	-
Maaskant et al. (2016) [33]^a	COMFORT	Poor, Fair, Good, and excellent*	Poor, Fair, Good, and excellent*	-	Yes	-

Author (year)	Validity assessment				Clinical utility/feasibility
	Discriminant validity	Cross-cultural validity	Criterion/ concurrent validity	Construct / Structural/ validity	
Chan, et al. (2022) [24]^a	-	-	Low	Moderate	Low responsiveness
	-	-	High	-	Moderate responsiveness
	-	-	Low	-	Moderate responsiveness
	-	-	Very low	-	-
	-	High	Moderate	-	Low responsiveness
Bai & Jiang. (2015) [25]^b	Good	Supported	Supported	Acceptable to high	Very good feasibility
	-	Supported	-	Acceptable to high	Very good feasibility
	-	Supported	Supported	Acceptable to high	Very good feasibility
Crellin, et al. (2007) [26]^a	-	-	Moderate to strong (r=0.743–0.921)	Weak to good	Burdensome to apply
	Demonstrated	-	Good	Supported	-
	Demonstrated	-	-	Good	-
	-	-	-	Demonstrated	-
Crellin, et al. (2015) [27]^a	Low	Supported	Low	-	Low to moderate feasibility and utility
Crellin, et al. (2018) [28]^a	Nil, Low, Moderate, High	Not reported	Nil to Low	-	Insufficient evidence of feasibility or clinical utility
Crosta, et al. (2014) [29]^(d)	-	-	Supported	Supported	Limited feasibility and clinical utility
Duhn & Medves. (2004) [30]^a	Established	-	-	-	None reported
	Established	-	-	-	None reported
	-	-	-	-	Low
	-	-	-	Supported (r=0.85, P<0.05))	Good
	Demonstrated	-	-	-	Limited
	Weak	-	-	-	None reported
	-	-	Established (With VAS) r=0.68 and 0.74. p<0.001	Established	None reported
	-	-	Good, (with TPPPS) Agreement 87.4%	-	Easy to use

Table 4 (continued)

Author (year)	Validity assessment				Clinical utility/ feasibility
	Discriminant validity	Cross-cultural validity	Criterion/ concurrent validity	Construct / Structural/ validity	
Giordano et al. (2019) [31] ^a	-	-	Established (with WBFPS)	Established	None reported
	-	-	Established (with VAS)	Established	
	-	-	Established (with CHEOPS)	Established	
	-	-	-	Established	
	-	-	Established (with TPPPS)	Established	
	Yes	-	Established (with VAS)	Established	
	-	-	Established (with CASS)	Established	
	-	-	Established (with EDIN, FLACC, TPPPS, CHEOPS)	-	
	-	-	Established (with MBPS, NIPS, OPS)	established	
	-	-	Established (with PASS)	-	
	-	-	Established (with COMFORT, VAS)	-	
	-	-	Established (with FLACC, VAS)	Established	
	-	-	Established (with, VAS)	Established	
	-	-	Established (with, NFCSr)	Established	
	-	-	-	Established	
	-	-	Established (with, MBPSr)	Established	
	-	-	Established (with, FLACC, NIPS)	-	
	-	-	-	Established	
	-	-	Established (with, MAPS, NIPS)	-	
	-	-	-	Established	
	-	-	-	Established	
	-	-	-	Established	
	-	-	-	Established	
	-	-	-	Established	
	-	-	-	Established	
	-	-	Established (with VAS)	Established	
	-	-	Established (with NRS)		
	-	-	Established (with BWFS)	Established	

Table 4 (continued)

Author (year)	Validity assessment				Clinical utility/feasibility
	Discriminant validity	Cross-cultural validity	Criterion/ concurrent validity	Construct / Structural/ validity	
Kingsnorth et al. (2015) [34] ^f	Well-established				Moderate clinical utility
Stapelkamp et al. (2011) [32] ^c	-	-		Demonstrated	High usability, Moderate clinical utility
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
	-	-	-	Demonstrated	
Maaskant et al. (2016) [33] ^a	Demonstrated	-	-	Poor, Fair and Good*	Not reported, Fair, Good, and Excellent for responsiveness

Acceptable (Inter-rater reliability) = $r \geq 0.60$

a Cronbach alpha, *K* Kappa, *ICC* Intraclass Correlation Coefficient, *r* correlation coefficient (Pearson's *r*), *CVI* Content Validity Index, - No information provided, *AHTPS* Alder Hey Triage Pain Scale, *BOPS* Behavioural Observational Pain Scale, *BPS* Behavioural Pain Scale, *CAAS* Cardiac Analgesic Assessment Scale, *CHEOPS* The Children Hospital of Eastern Ontario Pain Scale, *CHIPPS* Children's and Infants Postoperative Pain Scale, *CHIPPS or KUSS* Children's & Infant's Postoperative Pain Scale, *CSS* Clinical Scoring System, *CMPPMS* Chedoke-McMaster Paediatric Pain Management Sheet, *DEGR* Douleur Enfant Gustave Roussy Scale, *DCHPT* Derbyshire Children's Hospital Pain Tool, *FLACC* Face, Legs, Activity, Cry and Consolability, *EVENDOL* Evaluation Enfant Douleur, *FPS-R* Faces Pain Scale Revised, *MAPS* Multidimensional Assessment Of Pain Scale, *MAX* Maximally Discriminative Facial Movement Coding System, *MBPS* Modified Behavioural Pain Scale, *MIPS* Modified Infant Pain Scale, *mo* Months, *NAPI* Nursing Assessment of Pain Intensity Revised, *NFCSr* Neonatal Facial Coding System-Revised, *NIPS* Neonatal Infant Pain Scale, *ObsVAS* Observer Visual Analog Scale, *PEPPS* Preverbal, Early verbal Pediatric Pain Scale, *P-MIPS* Partial Modified Infant Pain Scale, *POCIS* Pain Observation Scale for Young Children, *POPS* Postoperative Pain Score, *PPP* Paediatric Pain Profile, *PRS* Pain Rating Scale, *RCEMCPS* Royal College of Emergency Medicine Composite Pain Scale, *RIPS* Riley Infant Pain Scale, *TPPPS* Toddler Pre-schooler Postoperative Pain Score, *TVP* Touch Visual Scale; *UWCH*: University of Wisconsin Children's Hospital Pain Scal

^a Rating based on the COSMIN checklist quality

^b Rating based on the Psychometric property coding system- CODE

^c Rating based on the SIGN-grading system

^d Rating system not described

^e Rating based on Risk of bias_ Combined ROBIS domains

^f Rating system not described, except for overall evaluation of V&R and clinical utility using rating schema for assessing clinical utility

* statistically significant difference

Evaluation ENfant DOuLeur (EVENDOL) [31]. For the MBPS, inconsistent level of evidence was reported [28].

Reliability

Reliability assessment of the pain tools was evaluated using internal consistency, inter-rater and/or intra-rater reliability. There were no reliability data reported for three of the tools, namely the Behavioral Pain Score (BPS), Partial Modified Infant Pain Scale (MIPS), and Touch Visual Pain scale (TVP). Several reviews reported that tools had internal consistency or inter/intra-rater reliability without providing details on type of reliability

data obtained for the tool [24, 31, 32, 34]. Although, in the case of Stapelkamp et al. [32] acceptable inter-rater reliability (correlation ≥ 0.6) was used as an inclusion criterion for observer-rated tools, without additional data being provided for individual tools in the review.

In six reviews the authors presented an evaluation of the methodological quality of the studies used to evidence the reliability of a range of tools including the FLACC, Toddler Pre-schooler Postoperative Pain Score (TPPPS), Preverbal, Early verbal Pediatric Pain Scale (PEPPS), MBPS, COMFORT-B [24, 26–28, 33]. The quality of the studies varied considerably from poor, poor/fair

to good/excellent. Issues around the sample size, lack of blinding, assessments using comparator tools completed by the same person, and the comparator not being an acceptable standard were often cited as reasons for quality issues in the primary studies. When the evaluation of quality was applied to the findings of the primary studies to derive the level of evidence of reliability this too varied from low to moderate to high for the FLACC [27] and MBPS [28].

Studies used Cronbach alpha for reports of internal consistency, whereas kappa coefficients, percentage agreement, correlation coefficients, and intra-class correlation coefficients were used to describe inter and intra-rater reliability. The variation in methods used for calculation of reliability and reports made direct comparisons difficult.

Validity

Validity assessment of the pain tools was reported using indicators of face/content validity, convergent validity, discriminant validity, criterion/concurrent validity, and/or construct validity. Evidence was provided in four reviews [24–27] on the cross-cultural validity of six assessment tools, namely the FLACC, COMFORT-B, POCIS, CHEOPS, PPP, and TPPPS, when used in a different language or cultural context. Additionally, some reviews reported that tools had ‘established’ or ‘supported’ criterion validity [31] and similarly so for discriminant validity [31, 34] without providing supportive data or their reason for this conclusion. In the case of Stapelkamp et al. [32] observational pain assessment tools were included if there was “both demonstrated known groups validity and inter-rater reliability.” In this case, demonstration of construct validity (i.e., the ability to differentiate between pain and non-pain states) was seen as more important than cross-validation of tools with others (criterion validity) in the absence of a gold standard. Data used to support the various tools construct validity was not provided within the review itself.

Studies evaluated validity of the pain assessment tools mostly reported ‘acceptable’ to ‘high’ quality [25–28, 33, 34]. One study identified weak or limited validity data for the BPS [30] and there were no validity data reported for two assessment tools, namely the Maximally Discriminative Facial Movement Coding System (MAX), and Royal College of Emergency Medicine Composite Pain Scale (RCCEMCP). Overall, the most common tools used when comparing pain scores of one assessment tool to another (criterion/concurrent validity) were the Observer administered Visual Analogue Scale (ObsVAS), Neonatal Infant Pain Scale (NIPS), FLACC and COMFORT.

Clinical utility and feasibility

Data on the utility, responsiveness and feasibility of pain assessment tools used in infants were very limited. Only eight tools (FLACC, COMFORT, POCIS, MAX, CSS, Children’s & Infant’s Postoperative Pain Scale (CHIPPS), CHEOPS, and PPP) were reported to have some evidence of utility and/or feasibility. Specific evaluation for feasibility appears to have been carried out only for three tools, namely the FLACC [24, 25], COMFORT [24, 25] and POCIS [25] whereas specific data on clinical utility have been undertaken for only one tool namely: the PPP [24, 29, 32, 34] with a limited focus for use in children with cognitive impairment.

An instance of conflicting data on utility and feasibility of the FLACC from two different reviews [25, 27] was attributed mainly to heterogeneity of studies, population and settings used to evaluate feasibility of the FLACC and therefore make it difficult to confidently draw broad conclusions on the tool’s feasibility and/or utility. It must be highlighted that when Bai and Jiang [25] assessed feasibility of the FLACC, COMFORT and POCIS, the main aspects assessed by the reviewers were the short length of items and clear user instructions. On the other hand, and for the PPP, earlier reviews clearly indicated that claims of tool feasibility and/or utility were based on time required to administer the tools [24, 29, 34] and/or brevity of a tool to its usefulness [24, 29, 32, 34]. Additional dimensions of clinical utility assessed by Kingsnorth et al. [34] included comprehensiveness, such as impact of pain, consideration of varied gross motor abilities, and considerations of varied verbal abilities of the child. For the PPP, another review suggested additional specifications for the tool suitability, such as age and cognitive abilities of the child, clinical setting, and the need for additional training to use by the health care practitioners [32].

The need for additional training/teaching in the use of the PPP was also judged as important by Crosta et al. [29], mainly with relation to the need for better instructions on scoring of the 20 items of the PPP to add clarity and help clinicians in the interpretation of the score’s meaning. For the PPP it was stated that authors of studies/tools did not report on what degree they found the PPP easy to use, or if it was time-consuming and complex for assessment and documentation, or the degree of friendliness for both child and family [32]. There were four tools which were described as ‘easy to use’ by multiple reviews, namely the CHEOPS, CHIPPS, CSS, and MAX, without providing further supporting evidence [30]. Crellin et al. [28] suggested there was insufficient data to evaluate the feasibility or utility of the MBPS.

Discussion

Eccleston et al. [7] have set making pain visible as one of their translational goals to improve pediatric pain management. To achieve this there is a need for valid and reliable observational pain assessment tools which inform clinical decision making and can be easily incorporated into clinical practice. Therefore, there is a need for evidence which supports the psychometric properties (validity and reliability) and feasibility as well as clinical utility of such tools, which is what this meta-review aimed to synthesize. What was demonstrated is that there are over 36 pain assessment tools which can be used to evaluate pain in infants aged 1 to 12 months. However, based on currently available data, no single tool can be recommended as the “gold standard”. This confirms the need for further research and development with the focus on arriving at a pain assessment tool which has sound psychometric and clinical utility properties. For our study we adopted the NICE definition as a gold standard (<https://www.nice.org.uk/glossary?letter=g>): “A method, procedure or measurement that is widely accepted as being the best available to test for or treat a disease”. In our case to achieve the status of gold standard it had to be considered the best available tool to measure pain in terms of reliability, validity, feasibility and usability.

Only one of the 11 reviews included in the current meta-review focused specifically on infants [30], the remainder included across a range of aged groups, such as premature and full-term neonates, toddlers, young children, and adolescents. From these reviews we extracted and analyzed data for 36 tools. The majority of the tools were multidimensional (exceptions being the NFCS-R, FPS-R, MAX and the ObVAS) and varied in their derivation, content, and rating of the pain scores. Many of the tools have been applied to children who are intubated, sedated and ventilated, including infants. Further, they have been used for acute pain, pre-, peri- and post-operative pain, procedural pain, including that associated with immunization, and chronic pain. An important lesson learnt that pain assessment tools in infants were mostly focused on acute pain, whereas chronic pain was relatively ignored. Of the tools identified in this review, the COMFORT scale was the main tool that can differentiate stages of pain and assess pain temporality. However, traditional definition of chronic pain may not apply in infant population. Therefore, our discussion in this regard has been hampered by difficulties in rectifying differences between acute and persistent/prolonged pain in definitions, as well as in their clinical picture. The fact that infants have not lived enough to meet chronic pain criteria applied to older children, and that most studies were designed to measure acute pain in predictable situations, such as painful procedures, may have contributed to the

researcher’s inability to predict or adequately measure chronic pain [38]. The use of explicit pain definitions and the integration of biological and behavioral/observational evidence could ultimately inform a more accurate assessment of infants who suffer from prolonged and persistent pain and assist in pain management.

The most cited tools in the included reviews were the FLACC, COMFORT, CHEOPS, followed by the MBPS, POPS, TPPPS and PPP. The findings of this meta-review in terms of the most commonly cited tools were in line with those identified by an earlier meta-review by Andersen et al. [11]. They reported that the previous systematic reviews had most frequently recommended the FLACC, COMFORT, and CHEOPS. Yet differences between the reviews in terms of their inclusion criteria, rating system and incomplete data synthesis made it impossible to recommend one particular tool to use. Given the variety of clinical settings in which the tools may be used and the different etiologies of pain this may not be unexpected. However, further evidence of the psychometric properties and clinical application and utility of these tools in practice would allow such recommendation to be made. Routine application of the COSMIN “Guideline for selecting outcome measurement instruments for outcomes included in a Core Outcome Set” [39] in the evaluation of current and future pain assessment tools would allow potential aggregating of data, thus strengthening the evidence pool.

This meta-review found that measures of inter-rater reliability, face/content validity and construct validity are the types of measurement properties most commonly and comprehensively reported, whereas information on clinical utility and cross-cultural validity is seldom and poorly reported. While it may be argued that certain aspects of validity and reliability are more important than others, we believe that evidence should be looked at as a whole. To date, there is no hierarchy of evidence described for the different psychometric properties described by earlier studies. Therefore, the collective opinion of earlier studies is regarded as the best available source of advice on offering an evaluation about what constitutes a validated and a reliable tool. Nevertheless, while many pain scales have focused on validity related to inter-rater and intra-rater reproducibility, it could be argued that looking at more scale properties, such as responsiveness to a known treatment may be more fundamental.

As suggested by Duhn & Medves [30] further evaluation of clinical utility and feasibility of existing pain assessment tools is needed to enhance our ability to accurately assess pain in the infant population. Furthermore, evaluation of clinical utility and feasibility would provide important information in regard to the appropriateness of existing tools across different clinical

settings. As suggested by Eccleston et al. [7] in clinical practice pain in infants tends to be both underdiagnosed and undertreated by clinicians. This is compounded by the lack of a gold standard tool used to measure pain in infants. Certainly, a gold standard cannot be achieved without the tool having a clearly evidenced clinical utility and feasibility, in addition to sound psychometric properties.

Studies focused on tools' clinical utility and feasibility generally have not moved forward in a systematic nor rapid way because of the logistical problems connected with use of pain assessment tools in vulnerable populations. It could also be related to the lack of clarity about what constitutes clinical utility in the literature and how it can be evaluated [40]. The length of items used in pain assessment tools, ease of use, time needed to complete/document the assessment, and clear user instructions are examples of aspects that have the potential to influence both feasibility of use and clinical utility in clinical practice. Therefore, these should be reported by studies, in addition to other evaluation of reliability and validity. Additionally, future studies should also consider current conflicting data on utility and feasibility of tools [24, 25, 27, 29, 32, 34] by addressing issues arising from heterogeneity of studies, dimensions evaluated, heterogeneity of populations and settings that makes it difficult to conclusively evaluate clinical utility and feasibility.

It is worth noting that Patient Related Outcomes Measures (PROMs) [39] are critical to evidencing the clinical utility of pain assessment tools. Establishing valid and reliable assessment of pain in clinical practice leads to improved patient outcomes (e.g., less pain, better pain management, better quality of life). Proctor's conceptual model of implementation science outcomes is another approach that could be used in future studies to present evidence of the feasibility of use of current pain assessment tools in infants and related outcomes [41]. The described implementation science methodology and the PROMs could be used as a framework for future studies.

We noted also that nine of the 11 reviews were conducted in Western countries and cultural validity of the pain assessment tools were only addressed in four; one review by Bai & Jiang [25] with special focus Chinese children and included the FLACC, COMFORT-B and POCIS, and those of Crellin et al. [24, 26–28] which covered the FLACC, TPPPS and MBPS. There are several variations in the way different individuals interpret pain assessment tools as a function of their age, literacy levels, and cultural background [42]. We can argue, therefore, that the appropriateness of using one measure created in one culture or setting for another setting is a matter of concern, especially in developing countries where the social, educational, and economic context differs

dramatically from Western countries. Exploration of cross-cultural validity provides suggestive evidence of potential for use within diverse cultural and linguistic context.

Given that there is still no ground truth for the measurement of pain in infants, certain aspects of validity testing are hard to estimate. COSMIN guideline[35] for systematic reviews of patient-reported outcome measures is increasingly seen as measurements that may help overcome of the above-mentioned methodological problems. Future studies can achieve greater rigor and relevance by (i) ensuring that a tool's 'responsiveness' is tested against known analgesics; (ii) testing 'construct validity' such as 'hypothesis testing validity' and (iii) ensuring that the Importance/Assessment of 'Interpretability' is considered when constructing and testing new tools.

Considering difficulties in assessing pain in populations unable to communicate, including infants, parental involvement remains an important component for assessment of pain. Nurses and clinicians often rely on parent's input to document infant's behavioral responses, especially when the child is too young to verbalize their pain or symptoms[43]. Inclusion of providers and parental input to identify the early stages of pain is similar to that made previously by an earlier review[38] and is based on pediatric committee consensus [43]. There is also a growing interest in addressing such challenges by leveraging from technological advances. In this regard, a review by Zamzmi et al. [3] showed that the use of automated methods for pain analysis and recognition may be a better fit for pain assessment in infants. Therefore, considering developments in machine learning based techniques for measuring pain in this population group, it can be expected that automation of pain assessment could have a number of implications in clinical practice that will result in a) development of pain assessment tools whose automation is driven by artificial intelligence; b) digitization of pain assessment documentation; c) enabling of rapid point-of-care assessment and d) continuous monitoring of pain. Continuous monitoring and predictive algorithms developed by artificial intelligence (AI) is another approach that may assist in measuring aspects of pain, including chronic pain[42]. These developments may well have a positive impact on addressing the four transformative goals proposed by Eccleston et al. [7] and therefore improve the lives of children experiencing pain as well as their families.

Strengths and limitations

The methodology approach adopted for this meta-review was a strength as it provided a comprehensive and a practical way of obtaining information on currently used

pain assessment tools compared to undertaking reviews of primary studies. Another strength is the rigorous approach adopted in the review process as we attempted to minimize bias by having each review screened independently by at least two reviewers. The main limitation of this study was that identification of all pain assessment tools used in infants could not be guaranteed, as not all of them may have been included in identified systematic reviews. Furthermore, pain assessment tools not published in English may also have been missed. Whilst we checked further available data when readily available, such as in supplementary files, we did not communicate with the authors to obtain missing data. Finally, our search did not specifically target reviews reporting on the feasibility or clinical utility of the pain assessment tools, therefore our evidence may be more limited on these aspects than it might have been otherwise.

Conclusion

The assessment of pain in infants remains complex and challenging for practitioners and researchers, mainly because infants are unable to self-report their pain. The 11 published reviews included in this meta-review evaluated 36 tools of which seven of these were included at least by three reviews. The level of evidence reported in regard to each tool's psychometric properties varied widely across published reviews, and only a few of these reviews reported on feasibility and clinical utility and had conflicting results or measured different dimensions of utility. Therefore, we were unable to conclude that any one tool is better than the other without proper caution. To address this there is a need for standardization of the evaluation of psychometric properties of pain assessment tools, together with a greater focus feasibility and clinical utility. Underpinning the latter is the need for accompanying PROMs data, indicating the systematic use of the tool results the best possible pain management outcomes.

Abbreviations

AHTPS	Alder Hey Triage Pain Scale
BOPS	Behavioural Observational Pain Scale
BPS	Behavioural Pain Score
CAAS	Cardiac Analgesic Assessment Scale
CHEOPS	The Children Hospital of Eastern Ontario Pain Scale
CHIPPS	Children's and Infants Postoperative Pain Scale
CHIPPS or KUSS	Children's & Infant's Postoperative Pain Scale
CSS	Clinical Scoring System
CMPPMS	Chedoke-McMaster Paediatric Pain Management Sheet
DCHPT	Derbyshire Children's Hospital Pain Tool
FLACC	Face, Legs, Activity, Cry and Consolability
EVENDOL	Evaluation Enfant Douleur FPS-R: Faces Pain Scale Revised

MAPS	Multidimensional Assessment Of Pain Scale
MAX	Maximally Discriminative Facial Movement Coding System
MBPS	Modified Behavioural Pain Scale
MIPS	Modified Infant Pain Scale
Mo	months
NAPI	Nursing Assessment of Pain Intensity Revised
NFCSr	Neonatal Facial Coding System-Revised
NIPS	Neonatal Infant Pain Scale
ObsVAS	Observer Visual Analog Scale
PEPPS	Preverbal, Early verbal Pediatric Pain Scale
P-MIPS	Partial Modified Infant Pain Scale
POCIS	Pain Observation Scale for Young Children
POPS	Postoperative Pain Score
PPP: Paediatric Pain Profile; PRS	Pain Rating Scale
RCEMCPS	Royal College of Emergency Medicine Composite Pain Scale
RIPS	Riley Infant Pain Scale
TPPPS	Toddler Pre-schooler Postoperative Pain Score
TVP	Touch Visual Scale
UWCH	University of Wisconsin Children's Hospital Pain Scale

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12887-023-04099-7>.

Additionalfile1:SupplementaryTable1. Selection Criteria for the Systematic Review (PICOS). **SupplementaryTable2.** Databased searched, search dates and outputs. **SupplementaryTable3.** Database search strategies and outputs. **SupplementaryTable4.** List of search terms. **SupplementaryTable5.** List of Excluded reviews. **SupplementaryTable6.** List of included reviews.

Acknowledgements

Not applicable.

Authors' contributions

DA, EM, JH & KH conceived and designed the study. JH and KH performed search. DA and EM screened data and performed the analysis. DA and EM performed critical appraisal of studies, DA drafted the manuscript and all authors assisted in editing and writing of the manuscript. All authors gave final approval of this version to be published.

Funding

This project was funded by Paincheck and fund received for this project has been paid to ECU to cover the salary award of the research assistants.

Availability of data and materials

All data generated or analysed during this study are included in this published article [and its [supplementary information files](#)].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

JDH was employed by PainChek Ltd as its Chief Scientific Officer in February 2020, previous to that JDH worked under a research contract between Curtin University and PainChek Ltd on the development and evaluation of PainChek's Adult and Infant apps (formerly known as EPAT Technologies).

Received: 14 September 2022 Accepted: 28 May 2023
Published online: 19 June 2023

References

- Miller-Hoover SR. Using Valid and Reliable Tools for Pain and Sedation Assessment in Pediatric Patients. *Crit Care Nur.* 2019;39(3):59–66. <https://doi.org/10.4037/ccn2019713>.
- Olsson E, Ahl H, Bengtsson K, Vejayaram DN, Norman E, Bruschetti M, Eriksson M. The use and reporting of neonatal pain scales: a systematic review of randomized trials. *Pain.* 2021;162(2):353. <https://doi.org/10.1097/j.pain.0000000000002046>.
- Zamzmi G, Kasturi R, Goldof D, Zhi R, Ashmeade T, Sun Y. A review of automated pain assessment in infants: features, classification tasks, and databases. *IEEE Rev in Biomed Eng.* 2017;11:77–96. <https://doi.org/10.1109/RBME.2017.2777907>.
- Mathew PJ, Mathew JL. Assessment and management of pain in infants. *Postgrad Med J.* 2013;79(934):438–43. <https://doi.org/10.1136/pmj.79.934.438>.
- Talbot K, Madden VJ, Jones SL, Moseley GL. The sensory and affective components of pain: are they differentially modifiable dimensions or inseparable aspects of a unitary experience? A systematic review. *Brit J Anaesth.* 2019;123(2):e263–72. <https://doi.org/10.1016/j.bja.2019.03.033>.
- Herr K, Coyne PJ, Ely E, Gélinas C, Manworren RC. Pain assessment in the patient unable to self-report: clinical practice recommendations in support of the ASPMN 2019 position statement. *Pain Manag Nurs.* 2019;20(5):404–17. <https://doi.org/10.1016/j.pmn.2019.07.005>.
- Eccleston C, Fisher E, Howard RF, Slater R, Forgeron P, Palermo TM, Wood C. Delivering transformative action in paediatric pain: a Lancet Child & Adolescent Health Commission. *Lancet Child Adolesc Health.* 2021;5(1):47–87.
- Beltramini A, Milojevic K, Pateron D. Pain assessment in newborns, infants, and children. *Ped Annals.* 2017;46(10):e387–95. <https://doi.org/10.3928/19382359-20170921-03>.
- Kahsay, H. Assessment and treatment of pain in pediatric patients. *Current Pediatric Research.* Int J Pediatr. 2017. <https://www.currentpediatrics.com/article/assessment-and-treatment-of-pain-in-pediatric-apteints.html>.
- Attallah-Wasif ES. Pain in infants, children, and adolescents. In: Abd-Elsayed A, editor. *Pain.* Cham: Springer; 2019. https://doi.org/10.1007/978-3-319-99124-5_236.
- Andersen RD, Langius-Eklöf A, Nakstad B, Bernklev T, Jylli L. The measurement properties of pediatric observational pain scales: A systematic review of reviews. *Int J Nur Stud.* 2017;73:93–101. <https://doi.org/10.1016/j.ijnurstu.2017.05.010>.
- Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *JBI Evidence Implementation.* 2015;13(3):132–40. <https://doi.org/10.1097/XEB.0000000000000055>.
- Munn Z, Aromataris E, Tufanaru C, Stern C, Porritt K, Farrow J, Jordan Z. The development of software to support multiple systematic review types: the Joanna Briggs Institute System for the Unified Management, Assessment and Review of Information (JBI SUMARI). *JBI Evidence Implementation.* 2019;17(1):36–43. <https://doi.org/10.1097/XEB.0000000000000152>.
- Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, Koffel JB. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews.* 2021;10(1):1–19. <https://doi.org/10.1186/s13643-020-01542-z>.
- World Health Organization (Western Pacific). Newborn health. WHO, 2021. Retrieved from <https://www.who.int/westernpacific/health-topics/newborn-health>.
- Raja SN, Carr DB, Cohen M, Finnerup NB, Flor H, Gibson S, Vader K. The Revised International Association for the Study of Pain definition of pain: Concepts, challenges, and compromises. *Pain.* 2020;161(9):1976–82.
- Stevens BJ, Hathway G, Zempsky WT. *Oxford Textbook of Pediatric Pain.* Glasgow: Oxford University Press; 2021. <https://doi.org/10.1093/med/97801988187620001.0001>.
- Lu G, Li X, Li H. Facial expression recognition for neonatal pain assessment. *Nanjing: International Conference on Neural Networks and Signal Processing;* 2008. p. 456–60. <https://doi.org/10.1109/ICNNSP.2008.4590392>.
- Johnston CC, Stevens B, Craig KD, Grunau RV. Developmental changes in pain expression in premature, full-term, two-and four-month-old infants. *Pain.* 1993;52(2):201–8. [https://doi.org/10.1016/0304-3959\(93\)90132-9](https://doi.org/10.1016/0304-3959(93)90132-9).
- Johnston CC, Stevens BJ, Yang F, Horton L. Differential response to pain by very premature neonates. *Pain.* 1995;61(3):471–9. [https://doi.org/10.1016/0304-3959\(94\)00213-X](https://doi.org/10.1016/0304-3959(94)00213-X).
- Johnson N, Phillips M. Rayyan for systematic reviews. *J Electron Resour Librarianship.* 2018;30(1):46–8. <https://doi.org/10.1080/1941126X.2018.1444339>.
- Walsh WB, & Betz NE. *Tests and assessment.* Prentice-Hall, Inc 1995.
- Aromataris E, Pearson A. The systematic review: an overview. *AJN The Am J Nurs.* 2014;114(3):53–8. <https://doi.org/10.1097/01.NAJ.000044496.24228.2c>.
- Chan AY, Ge M, Harrop E, Johnson M, Oulton K, Skene SS, Wong IC, Johnsson L, Howard RF, Liossi C. Pain assessment tools in paediatric palliative care: A systematic review of psychometric properties and recommendations for clinical practice. *Palliative Medicine.* 2022;36(1):30–43. <https://doi.org/10.1177/02692163211049309>.
- Bai J, Jiang N. Where are we: A systematic evaluation of the psychometric properties of pain assessment scales for use in Chinese children. *Pain Manag Nur.* 2015;16(4):617–31. <https://doi.org/10.1016/j.pmn.2014.11.003>.
- Crellin D, Sullivan TP, Babl FE, O'Sullivan RONAN, Hutchinson A. Analysis of the validation of existing behavioral pain and distress scales for use in the procedural setting. *Pediatric Anesthesia.* 2007;17(8):720–33. <https://doi.org/10.1111/j.1460-9592.2007.02218.x>.
- Crellin DJ, Harrison D, Santamaria N, Babl FE. Systematic review of the Face, Legs, Activity, Cry and Consolability scale for assessing pain in infants and children: is it reliable, valid, and feasible for use? *Pain.* 2015;156(11):2132–51. <https://doi.org/10.1097/j.pain.0000000000000305>.
- Crellin DJ, Babl FE, Santamaria N, Harrison D. A systematic review of the psychometric properties of the Modified Behavioral Pain Scale (MBPS). *J of Ped Nur.* 2018;40:14–26. <https://doi.org/10.1016/j.pedn.2018.02.005>.
- Crosta QR, Ward TM, Walker AJ, Peters LM. A review of pain measures for hospitalized children with cognitive impairment. *J Spec Pediatric Nurs.* 2014;19(2):109–18. <https://doi.org/10.1111/jspn.12069>.
- Duhn LJ, Medves JM. A systematic integrative review of infant pain assessment tools. *Adv Neonatal Care.* 2004;4(3):126–40. <https://doi.org/10.1016/j.adnc.2004.04.005>.
- Giordano V, Edobor J, Deindl P, Wildner B, Goeral K, Steinbauer P, Olischar M. Pain and sedation scales for neonatal and pediatric patients in a preverbal stage of development: a systematic review. *JAMA pediatrics.* 2019;173(12):1186–97. <https://doi.org/10.1001/jamapediatrics.2019.335>.
- Stapelkamp C, Carter B, Gordon J, Watts C. Assessment of acute pain in children: development of evidence-based guidelines. *Int J Evid Based Healthc.* 2011;9(1):39–50. <https://doi.org/10.1111/j.1744-1609.2010.00199.x>.
- Maaskant J, Raymakers-Janssen P, Veldhoen E, Ista E, Lucas C, Vermeulen H. The clinimetric properties of the COMFORT scale: A systematic review. *Eur J Pain.* 2016;20(10):1587–611. <https://doi.org/10.1002/ejp.880>.
- Kingsnorth S, Orava T, Provvidenza C, Adler E, Ami N, Gresley-Jones T, Fehlings D. Chronic pain assessment tools for cerebral palsy: a systematic review. *Pediatrics.* 2015;136(4):e947–60. <https://doi.org/10.1542/peds.2015-0273>.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, De Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010;19(4):539–49. <https://doi.org/10.1007/s1136-010-9606-8>.
- Harbour R, Miller J. A new system for grading recommendations in evidence-based guidelines. *BMJ.* 2001;323(7308):334–6. <https://doi.org/10.1136/bmj.323.7308.334>.
- Zwakhalen SM, Hamers JP, Abu-Saad HH, Berger MP. Pain in elderly people with severe dementia: a systematic review of behavioural pain assessment tools. *BMC Geriatrics.* 2006;6(1):1–15. <https://doi.org/10.1186/1471-2318-6-3>.

38. Llerena A, Tran K, Choudhary D, Hausmann J, Goldgof D, Sun Y, Prescott SM. Neonatal pain assessment: Do we have the right tools? *Frontiers in Pediatrics*. 2022;10. <https://doi.org/10.3389/fped.2022.1022751>
39. Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., ... & Terwee, C. B. Guideline for selecting outcome measurement instruments for outcomes included in a Core Outcome Set. The Netherlands: COMET COSMIN.2016. Microsoft Word - 13063_2016_1555_MOESM2_ESM_Prinsen_track changes accepted_R.1 (ksu.edu.sa)
40. Smart A. A multi-dimensional model of clinical utility. *Int J Q Health Care*. 2006;18(5):377–82. <https://doi.org/10.1093/intqhc/mzl034>.
41. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Hensley M. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Mental Health Ment Health Serv Res*. 2011;38(2):65–76. <https://doi.org/10.1007/s10488-010-0319-7>.
42. Pathak, A., Sharma, S., & Jensen, M. P. The utility and validity of pain intensity rating scales for use in developing countries. *Pain reports*. 2018; 3(5). <https://doi.org/10.1097/PR9.0000000000000672>
43. Ghai S, Hakim M, Dannenbaum E, Lamontagne A. Prevalence of vestibular dysfunction in children with neurological disabilities: a systematic review. *Front Neurol*. 2019;17(10):1294. <https://doi.org/10.3389/fneur.2019.01294>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

